DOCUMENT RESUME

ED 257 821 SP 026 219

AUTHOR Fielding, Glen D.; Schalock, H. Del
TITLE Integrating Teaching and Testing: A Handbook for High School Teachers.
INSTITUTION Oregon State System of Higher Education, Monmouth. Teaching Research Div.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jan 85
CONTRACT 400-82-0013
NOTE 162p.
PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS *Instructional Improvement; *Learning Strategies; *Secondary Education; Skill Analysis; Student Evaluation; Teaching Methods; *Test Construction; *Testing

ABSTRACT

The intent of this handbook is to illustrate the relation between teaching and testing and to demonstrate how they can promote student learning. The first section, "Foundations," offers a discussion of broad ideas about teaching, learning, and testing, and their interrelationships. Chapter one describes generally the kinds of things that teachers do in classrooms where teaching and testing are integrated. Chapter two discusses essential aspects of learning which mutually affect instruction and testing. The third chapter deals with how different dimensions of the context of instruction (the types of students with whom the teachers is working, resources and support available, and the nature of the instructional models in use) influence teaching and testing practices. Chapter four describes the varied purposes served by tests, such as assessing students' past learning or their progress in learning. The fifth chapter offers an overview of the kinds of tests addressed in the handbook: objective tests, essays, and observation of performance. The second section, "Applications," focuses on tasks that need to be performed to integrate teaching and testing. Chapters are included on matching teaching and testing to desired outcomes, assuring quality in tests, preparing, administering, and scoring tests, and using test information for various instruction-related purposes. Lists of references and related resources are provided. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

December 1983
Revised, August 1984
Revised, January 1985

SCOPE OF INTEREST NOTICE

The Eric Facility has assigned this document for proccesing to:	SP
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should refect their special points of view	TM

ED257821

INTEGRATING TEACHING AND TESTING:

A HANDBOOK FOR HIGH SCHOOL TEACHERS

Developed by

Glen D. Fielding

H. Del Schalock

The Teaching Research Division
Oregon State System of Higher Education

through contract No. 400-83-0013 with the
National Institute of Education

SP026219

The following individuals contributed
to the development of this handbook:

Jason Millman (Consultant to the project),
Cornell University

Lewis Pike and Susan Klein (Project officers),
National Institute of Education

Technical Panel on Teaching

David Merrill (Chair)
University of Southern California

Jeri Benson
University of Maryland

Harvey Black
Brigham Young University

Bennie Lowery
San Diego School District

Charles Reigeluth
Syracuse University

Teacher Advisory Panel

Alice Befus
Willamina HS, OR

Jim Ferguson
Cascade HS, OR

Dave Hagfeldt
South Albany HS, OR

Jan Jobe
Central HS, OR

Diane John
Parkrose HS, OR

Darrell Johnston
Dayton HS, OR

Larry Johnston
Woodburn HS, OR

Bob Kenyon
Dallas HS, OR

Al Ko
Ft. Vancouver HS, WA

Jon Lewis
McMinnville HS, OR

Mike McVay
Glencoe HS, OR

Lou Mueller
Sunset HS, OR

Technical Panel on Testing

Ronald Hambleton (Chair)
University of Massachusetts

Eva Baker
University of California
Los Angeles

Ronald Berk
Johns Hopkins University

Tom Haladyna
Americal College
Testing Program

Joan Herman
University of California
Los Angeles

Paul Williams
Maryland Department
of Education

INTEGRATING TEACHING AND TESTING:

A Handbook for High School Teachers

Page

Introduction 1

PART I: FOUNDATIONS 3

Chapter 1: What does it mean to integrate teaching and testing? 5

Chapter 2: Dimensions of learning that influence teaching and testing 13

Chapter 3: Dimensions of context that influence teaching and testing 23

Chapter 4: Purposes served by tests 35

Chapter 5: Types of test items and assessment procedures 41

PART II: APPLICATIONS 53

Chapter 6: Formulating expected learning outcomes 55

Chapter 7: Matching teaching to expected learning outcomes 63

Chapter 8: Matching test items and assessment procedures to expected learning outcomes 77

Chapter 9: Assuring quality in tests items and assessment procedures 91

Chapter 10: Preparing, administering and scoring tests 115

Chapter 11: Using test information as feedback to students 135

Chapter 12: Using test information as a guide to instructional decisions 143

Chapter 13: Using test information in grading 159

INTRODUCTION

Why a Handbook for Integrating Teaching and Testing

Teaching and testing often appear distinct and separate from each other. It is not uncommon for teachers to see testing as a distraction from important work to be done, and for test specialists to regard teaching as outside their main sphere of interest. In virtually all colleges of education, curriculum and instruction are dealt with in one department, and measurement and testing are addressed in another. Teaching often is portrayed as an art or craft; testing frequently is characterized as a science or technology. A gap seems to exist in the eyes of many educators between instruction and assessment.

In this handbook the assumption is made that teaching and testing can be closely linked. The intent of the handbook is to illuminate the relation between these two processes and to demonstrate how they can be used together to promote student learning.

Of course, the notion that teaching and testing are mutually enhancing is not new. Many excellent books have been written on teaching methods and on classroom testing that suggest ways in which instruction and assessment are linked. However, texts that focus on teaching methods generally provide only a limited amount of information on assessment. In like manner, texts that focus on testing typically include only a brief discussion of instruction. Very few books or related materials have had as their central theme the teaching-testing connection. This handbook has been designed to bring this connection into sharper focus.

Content and Organization

The handbook is divided into two parts. The first, "Foundations," offers a discussion of broad ideas about "teaching," "learning" and "testing," and their interrelationships. The second, "Applications," furnishes guidelines for putting these ideas into practice.

The Foundations section consists of five chapters. Chapter 1 sets forth the premises that underlie the handbook, and describes generally the kinds of things that teachers do in classrooms where teaching and testing are integrated. Chapter 2 casts light on essential aspects of learning which mutually affect instruction and testing. Chapter 3 is an effort to show how different dimensions of the context of instruction, e.g., the types of students with whom a teacher is working, the kind of resources and support available to teachers, and the nature of the instructional models in use, influence teaching and testing practices. Chapter 4 furnishes a description of the varied purposes served by tests, like assessing students' past learning or their progress in learning during a term or semester. Chapter 5 offers an overview of kinds of tests and assessment procedures addressed in the handbook: objective tests; essays and other tasks in which students create products; and observation of performance.

The chapters included in the Applications section focus on the various tasks that need to be performed to integrate teaching and testing. The first chapter in this section deals with formulating expected learning outcomes and builds closely on Chapter 2 in the Foundations section. Chapters follow on matching teaching and testing to desired learning outcomes, on assuring quality in tests, on preparing, administering, and scoring tests, and on using test information for various instruction-related purposes.

At the end of each chapter a list of references and related resources is provided. The references indicate the sources used in preparing the chapter. Related resources include selected books and articles that the reader may wish to consult as a supplement to the information provided in the chapter.

Development of the Handbook

The handbook was developed and pilot tested through funds provided by the National Institute of Education, under the guidance of an advisory panel of 12 high school teachers from schools in western Oregon and southwestern Washington and two national advisory panels in the areas of instructional design and educational measurement. Members of these panels and other individuals who made important contributions to the handbook are identified on the inside of the front cover.

Intended Uses of the Handbook

The handbook is intended for high school teachers, although much of the content is appropriate for teachers in the lower grades as well.

It is anticipated that teachers generally will use this handbook in the context of a school-based staff development program on integrating teaching and testing. Such a program currently is being field tested, and its effects on teacher practice and student attitudes are being assessed. Guidelines and resources for conducting a staff development program of the kind being field tested will be available in the spring of 1985.

PART I

FOUNDATIONS

CHAPTER 1

WHAT DOES IT MEAN TO INTEGRATE TEACHING AND TESTING

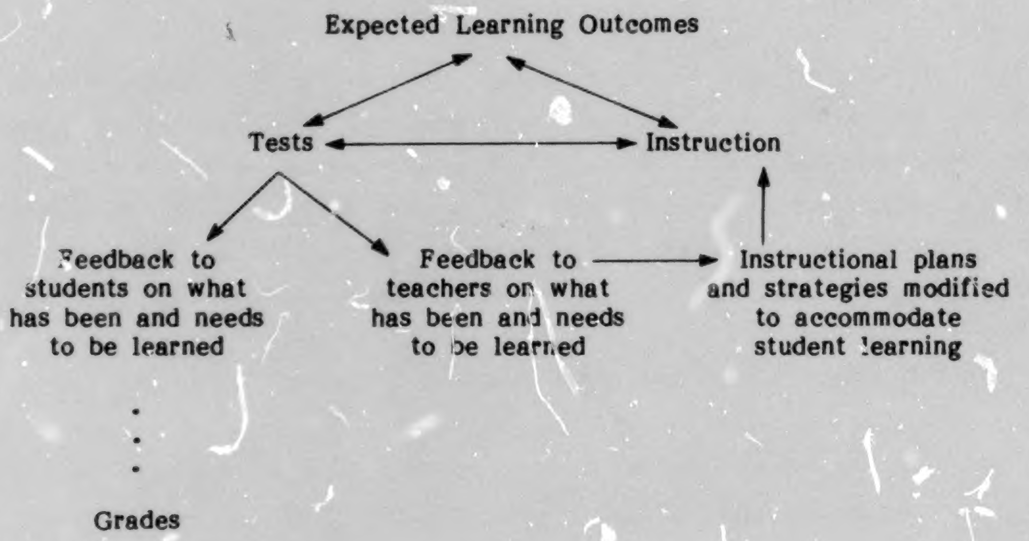
Introduction

The approach to teaching and testing set forth in this handbook rests on the assumption that both processes are influenced greatly by what and how well teachers expect students to learn. The position is taken that **expected learning outcomes** provide a basis for both the design of instruction and the development of tests, and a reference point against which test results can be reported, interpreted, and acted upon. The importance of learning goals to instruction and assessment is a theme running throughout the handbook.

Another assumption underlying the handbook is that the instructional **context** in which teachers operate influences teaching and testing practices. It is suggested that this context is composed of: (a) the kinds of students in a class; (b) the instructional models which the teacher or district has adopted; e.g., "Mastery Learning," or "Instructional Theory into Practice"; (c) the resources available for teaching and testing; and (d) the opportunities available for professional development. Although less attention is given in the handbook to accommodating teaching and testing to context than in anchoring these processes to expected learning outcomes, contextual factors are addressed explicitly in Chapter 3, and referred to in various sections of the chapters on formulating expected learning outcomes and on using test information for the variety of purposes it serves.

Expected Learning Outcomes as an Anchor for Teaching and Testing

The learning outcomes a teacher intends students to achieve carry direct implications for the selection of teaching strategies and materials, the construction of test items and related assessment procedures, and for the use of assessment information. The importance of learning goals in the teaching-testing process is illustrated below.


Expected Learning Outcomes
Tests
Instruction
Feedback to students on what has been and needs to be learned
Feedback to teachers on what has been and needs to be learned
Instructional plans and strategies modified to accommodate student learning
Grades

The components in the figure, and their interrelationships, are discussed in the paragraphs that follow.

Expected Learning Outcomes

The term, "expected learning outcome," refers to a desired result of the learning process, i.e., the knowledge, skills, or attitudes that students are expected to develop through a course of instruction.

In this handbook, it is recommended that statements of intended learning outcomes indicate the **content** students are to learn (facts, concepts, principles, or procedures), and **what students are to be able to do** in relation to this content (remember it; use it; or refine it).* The greater a teacher's clarity about the learning outcomes to be attained, the firmer is the basis for instruction and assessment.

---

* Although attitudes also represent a form of content, it was considered beyond the scope of the handbook to treat in depth the topic of attitude assessment. Resources related to this topic, however, are listed at the end of Chapter 3.

Instruction

Instructional strategies vary according to the type of learning outcomes desired. For example, strategies for teaching facts differ from strategies for teaching concepts. As another example, strategies for promoting students' knowledge of procedures differ from strategies for fostering their ability to apply a set of procedures. The nature of instructional presentations and the characteristics of learning activities depend to a large extent on the kind of learning outcomes students are expected to attain.

Tests

For purposes of this handbook, a test is viewed as a systematic procedure for measuring what students know or can do. Tests are considered to have three basic elements:

1. A clearly defined task that students are to perform;
2. Clearly defined conditions that govern students' performance; and
3. Clearly defined rules for scoring students' responses to the task.

While all tests must have these three basic elements, they can vary a great deal in **format.** Three kinds of formats are discussed in the handbook: (1) objective test items; (2) essay questions and other tasks in which products are produced, e.g., sculptures, mechanical drawings, laboratory reports; and (3) observation of performance.

Although each of these formats can be used to assess different kinds of learning outcomes, particular formats assess some types of outcomes more efficiently or directly than others. To cite a rather obvious example, observation of performance is not an especially well suited format for assessing students' knowledge of facts, but it is quite appropriate for assessing students' use of athletic or artistic techniques, processes of interpersonal communication, and certain problem solving strategies. Connections between test formats and types of expected learning outcomes are discussed in Chapter 8. Guidelines for constructing test items in each type of format are provided in Chapter 9.

There is more to building tests, however, than designing individual test items or other assessment tasks. Such items and tasks need to be organized to serve particular purposes, e.g., assessing what students have learned from a specific unit versus assessing what they have learned from a course as a whole, and in a manner that reflects the relative importance of each learning outcome to be assessed. In addition, attention needs to be given to the context in which tests are to be administered. (The context in which informal assessments are carried out, for instance, is quite different from the context in which formal tests are administered.) Finally, procedures for scoring tests need to be developed, including procedures for evaluating essays, other students' products, and "in-process" performance. Issues involved in test preparation, administration, and scoring are addressed in Chapter 10.

Feedback to Students

Assuming that learning goals have been communicated to students, and that tests have been prepared that correspond to them, students can use test information to check their progress toward accomplishing the learning outcomes expected. Feedback from tests that draws attention to specific learning goals or objectives that students have attained, that points out learning expectations that have yet to be reached, or that indicates what steps need to be taken to achieve the learning outcomes desired, is more valuable as an aid to learning than information that is highly general, for example, an overall test score.

A variety of feedback forms and procedures are discussed in Chapter 11. All of them depend in one way or another on a shared understanding between teacher and student about the learning outcomes to be achieved through a course of study.

Feedback to Teachers

Just as students can use test information to monitor and manage their learning, so teachers can use test results to evaluate and improve their instruction. In classrooms where teaching and testing are integrated, teachers adapt instruction in light of evidence on student performance.

When test results are tightly anchored to expected learning outcomes it is easier to assess their implications for instructional improvement than when test results lack a clear anchor to learning expectations. For example, a different instructional response is called for when test results reveal that students have difficulties in remembering concepts, principles, or procedures, than when results reveal that students can remember the content they have learned but cannot apply it.

In Chapter 12, guidelines are provided for dealing with common patterns of test performance, both for individual students and a class as a whole. These guidelines are intended to be useful in checking probable causes of unexpectedly low or high achievement, and in identifying options for responding to these causes.

Grades

It is recommended in the handbook that grades for a course be based on student achievement in relation to established learning goals. Basing grades on the attainment of desired learning outcomes is a challenging task, however, since a student may reach some learning goals and not others, or perform better on one measure of outcome attainment than another, e.g., a student may achieve a high score on an essay prepared outside of class, but a low score on an in-class test measuring achievement of the same goal. Guidelines for translating information on outcome attainment into grades are presented in the last chapter of the handbook.

Context as an Influence on Teaching and Testing

What and how teachers teach and test is shaped to a considerable extent by the context in which they operate. The types of students with whom a teacher is working, the kind of instructional models in use, the quality and extent of available resources, and opportunities for professional development influence basic aspects of teaching and testing.

Student Characteristics

Students' backgrounds, abilities, attitudes, and learning styles have implications for numerous teacher practices, including:

- the level of complexity and sophistication of learning outcomes expected;
- the kind of learning environment and learning activities that are created (Some students, for example, may need a much more structured environment than others); and
- the frequency with which tests are given (Low ability students may need to be tested after each learning step they take, whereas higher ability students may be capable of taking many steps independently without guidance from tests).

Student characteristics have a strong effect on teacher practices.

Instructional Models

The instructional models a teacher has adopted also affect patterns of teaching and testing. In recent years, a number of general, research-based models have been developed, including Mastery Learning (Block, 1974; Bloom, 1976); the Beginning Teacher Evaluation Study Model (Fisher et al., 1980); Active Teaching (Good & Grouws, 1979); and Instructional Theory into Practice (Hunter, 1976). In addition, more specialized, content-related models have been developed, for example, the Learning Cycles Model (Karplus & Others, 1980), which is grounded in Piaget's theory of cognitive development, and the Jurisprudential Model, which is used to teach secondary school students skills in analyzing public policy issues (Joyce & Weil, 1982; Oliver & Shaver, 1966). Each instructional model calls for specific teaching and assessment approaches. To illustrate the implications of different instructional models for instruction and assessment, a comparison of several models is presented in Chapter 3.

Resources

The resources that teachers are able to draw upon exert a powerful impact on the teaching-testing process. For example, the supply of instructional materials appropriate for different ability levels sets limits on a teacher's effectiveness in individualizing instruction. Tests or test item pools that correspond to the curriculum teachers are expected to implement, and technological advances like computer-assisted instructional programs and test scoring machines, also are important features of the instructional context. The type of professional tools available to teachers influence their approach to instruction and assessment, and their success in carrying it out.

Opportunities for Professional Development

Finally, the opportunities teachers have for upgrading their skills, developing new instructional materials or assessment tools, and interacting with colleagues around issues related to student learning influence the nature of instruction and assessment. For example, in districts that have provided inservice training in goal-based instructional strategies, or in using test information to evaluate and improve instruction, teachers may have a different approach to teaching and testing than they would in districts where they have not been provided these opportunities for professional development. A teacher's success in fostering student learning depends in many ways on the context in which he or she is working.

REFERENCES

Chapter 1

Block, J.H. (Ed.). Schools, society, and mastery learning. New York: Holt, Rinehart, and Winston, 1974.

Bloom, B.S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Fisher, C.W., Berliner, D.C., Filby, N.N., Marliave, R., Cohen, L.S. & Dishow, M.M. Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to Learn. Washington, D.C.: National Institute of Education, 1980.

Goo[illegible] & Grouws, D.A. The Missouri mathematics effectiveness project: an [illegible]mental study in fourth-grade classrooms. Journal of Educational Psychology, 1979, 71 (3), 355-362.

Hunter, M. Improved instruction. El Segundo, Ca.: TIP Publications, 1976

Joyce, B., & Weil, M. Models of teaching. Englewood Cliffs, N.J.: Prentice Hall, 1982.

Karplus, P. & Others. Science teaching and the development of reasoning. Berkeley, CA: Lawrence Hall of Science, University of California, 1980.

Oliver, D.W., & Shaver, J.P. Teaching public issues in the high school. Boston: Houghton Mifflin, 1966.

RELATED RESOURCES

Journal of Educational Measurement (special issue on linking achievement testing to instruction), 1983, 20, (2).

Rudman, H., Kelly, J., Wanous, D., Mehrens, W., Clark, C. & Porter, A. Integrating assessment with instruction: A review (1922-1980). East Lansing, Mi: Institute for Research on Teaching, Michigan State University, 1980.

Tyler, R. & White, S. (Eds.) Testing, teaching and learning: Report of a conference on research and testing. Washington, D.C.: U.S. Department of Education, National Institute of Education, 1979.

CHAPTER 2

DIMENSIONS OF LEARNING THAT INFLUENCE TEACHING AND TESTING

Introduction

The focus of this chapter is on three aspects of learning that influence teaching and testing: the **framework** used to organize learning goals; the **content** of learning; and the **level of performance** expected of students in relation to the content they are studying.

A **framework** for learning is the structure through which learning is organized. Two fundamental frameworks are identified in the handbook: the framework established by subject areas, like biology and English literature, and the framework represented by life-roles, like the role of consumer or citizen. Choice of a framework has important implications for the selection of learning goals.

The **content** of learning refers to the particular facts, concepts, principles and procedures students are studying. Different categories of content impose different requirements on teaching and testing.

The **level of performance** to be attained refers to what students are able to do in relation to content. Generally speaking, students are expected to remember content; to use it; or to refine it. Identifying the level of performance students are to achieve is a critical step in focusing instruction and assessment.

Toward the close of the chapter, relations between the dimensions of learning identified in this handbook and Bloom's well-known system for classifying educational objectives are examined. The reader is encouraged to consult Bloom's work as an additional resource for clarifying expectations for student learning.

Framework for Learning

Learning in grades 9-12 generally is structured by the requirements of subject areas or the requirements of life-roles. Subject areas include mathematics, English, science, history, government, foreign language, fine arts, physical education, and so forth. Life roles include the roles of family member, citizen, consumer, individual, and producer. Subject areas and life roles represent different frameworks within which to think about and organize learning.

Learning organized around life roles differs from learning organized around subject areas in at least two ways. One is that learning to carry out life roles generally involves the use of content from a variety of subject areas. For example, consider the problems that have developed in areas of the country in which large quantities of toxic waste have been deposited. Members of a community affected by toxic wastes might need scientific and medical knowledge to understand the possible effects of these wastes on people's health and on other forms of life. They may need knowledge of geography, sociology, and economics to evaluate plans for relocating families should the threat from toxic wastes prove persistent. They may need knowledge of government and politics to influence public policy on environmental protection. In cases like this, it is the ability to integrate knowledge and skills from a range of subject areas, rather than knowledge and skills in individual subject areas per se, that is of critical importance.

Learning anchored to life roles also differs from learning anchored to subject areas in the scope of knowledge and type of skills that are pertinent for study. Many of the ideas, sensitivities, and capacities needed to function effectively in the role of family member, for example, lie outside the boundaries of academic disciplines. Learning to perform in life roles not only involves learning to integrate content from various subject areas. It also involves developing different understandings and abilities than those usually associated with academic disciplines.

No assumption is made in the handbook that high school teachers should devote more or less time than they regularly do to instruction based on subject areas or life roles. What is stressed in the handbook is that the choice of a subject matter orientation or a life role orientation influences the way in which instruction is carried out and tests are designed.

Content of Learning

Academic disciplines and life roles provide two general frameworks for learning. One does not teach or assess a discipline as a whole, however. Nor does one teach to or assess performance in a life role as a whole. Instruction and testing require a more specific content focus.

Generally speaking, learning focuses on one of four categories of content:

- facts,
- concepts,
- principles, or
- procedures.

This is the case whether learning is structured by a subject area or by a life role (Merrill, 1983; Reigeluth, 1980).

Facts

Facts are pieces of information, such as a person's name, a date, the characteristics of a particular object, the details of an event, etc. Examples of facts include:

- Thomas Jefferson was the third president of the United States;
- The moon is 380,000 kilometers from the earth; and
- Christmas is celebrated on December 25th.

The importance of a fact is determined by its relation to other, broader elements of content, such as concepts and principles, or in relation to the requirements of a particular role. For example, in a history class that is studying the principles underlying the United States Constitution, facts about the characteristics and effects of the Articles of Confederation and about the political and philosophical perspectives of the framers of the Constitution might be pertinent. In preparing for the role of parents, facts about common children's illnesses and their remedies, or about the different needs children have at different stages of development, might become a focus of learning. Facts rarely are important as ends in themselves. Their importance derives from their connection to other aspects of knowledge, or to the information needs associated with a particular role or situation.

Concepts

Concepts are groups of objects, events, or symbols which all share a common characteristic and which are identified by the same name (Merrill, 1983, p. 8).

Each word in a language represents a concept. Concepts are therefore basic to understanding all fields of study and all life roles. Important concepts associated with two academic disciplines and two life roles are presented below.

SUBJECT AREA		LIFE ROLE	
Physics	**Sociology**	**Consumer**	**Family Member**
mass	norm	warranty	love
energy	role	compound interest	spouse
volume	status	trade-in-value	children
density	institution	fraud	housework

Principles

Principles are ordered combinations of concepts. They are used to explain why certain things are as they are, to predict what will happen in particular situations, or to make evaluations.

Principles, like the concepts of which they are composed, are essential to understanding all subject areas and life roles. Principles associated with two subject areas and two life roles are presented below.

SUBJECT AREAS

Economics	**Physics**
In a competitive market, when demand for a product exceeds the supply, the price of the product rises.	Heat energy flows from an object of higher temperature to one of lower temperature until both are at the same temperature.

LIFE ROLES

Family Member	**Citizen**
Children commonly act toward others the way their parents act toward them.	All members of a community deserve respect, regardless of race, religion, or national origin.

Procedures

A procedure is a way of doing something. Procedures (or guidelines, techniques, strategies) indicate how to achieve a particular kind of goal, solve a particular class of problem, create a type of product, carry out a physical or mechanical activity, study a set of materials, or investigate a series of issues. Examples of procedures include:

- Rules for finding the latitude and longitude of a location;
- Guidelines for administering an intelligence test to young children;
- Strategies for playing "net" in a tennis game;
- Directions for polishing furniture; and
- Methods for determining the specific gravity of a substance.

Procedures may be intellectual or "cognitive" in nature, such as problem solving procedures or procedures for analyzing a work of literature. They may be interpersonally or group oriented, such as procedures for holding a meeting, responding to issues raised by a classmate or colleague in discussion, or procedures for communicating with an audience during a speech. Procedures may also guide performance on manual, mechanical or physical tasks, like using a typewriter, repairing a fender on a car, and dribbling a basketball.*

Some procedures conform to a tight, step-by-step structure. Others are more loosely structured and are formulated as general guidelines. For example, procedures that may be developed to enhance interpersonal communication are less amenable to step-by-step analysis than procedures for repairing cameras or using a word processing machine.

In the context of academic disciplines procedures generally relate to the process of producing, verifying, or refining "truth," i.e., advancing human understanding. They may also relate to creating or appreciating beauty, as in the fields of fine arts and music. Procedures from specialized fields are useful in carrying out many activities in real-life settings. But subject-matter linked procedures need to be supplemented by many other kinds of procedures to meet the demands of life roles. Academic disciplines provide little training in procedures associated with parenting, community organizing, shopping and budgeting, changing jobs or careers, making new friends, watching television selectively, and so forth. Procedures used in scholarship and other specialized pursuits represent only one of many sets of procedures necessary for adult living.

Level of Performance to be Attained

The level of performance to be attained refers to what students are to be able to do in relation to the facts, concepts, principles and procedures they are studying. In broad terms, students are expected to **remember** content, to **use** it, or to **refine** it.

To **remember** content is to be able to recall, recognize, or paraphrase it. For example, students might identify the steps involved in using a word processor or restate the definition of "homeostasis." Performance at the remember-level requires that the information to be remembered has been explicitly presented to students. For example, the task, "Explain three causes of the Civil War," would call for performance at the remember-level to the extent that the causes of the war already had been made clear to students. The task would require performance at a higher level if, for example, students had to construct an explanation based on an independent analysis of primary documents from the pre-war period.

---

* In the framework developed by Merrill, physical activities are not included within the category of procedures; procedures are viewed strictly as steps in a cognitive process. We include physical procedures here in recognition of the importance of such procedures in many high school courses. However, the distinction between a procedure that guides thought and one that guides physical action should be noted, since each type of procedure carries somewhat different implications for instruction and assessment.

When students are expected to perform at the remember-level the content to be remembered should be presented in a meaningful context. For example, teachers would not present the psychological principle, "Behavior is shaped by its consequences," as an isolated piece of knowledge, but would provide illustrations of how the principle could be used to explain situations with which the students were familiar. When effort is made to relate new information to what students already know, content to be remembered takes on meaning. This is in contrast to rote learning, in which students merely memorize content without having any sense of its significance.

Few teachers are content to foster achievement solely at the remember-level, however, even if the process of remembering builds on a foundation of meaningful learning. Teachers typically place emphasis on developing students' skills in **using** content.

When students attain concepts at the **use-level**, they not only can remember the definition and examples of a concept, but they can group unfamiliar objects, events, symbols, or ideas in terms of the concept. For example, instead of merely learning a stated definition of the concept "bias," students would be expected at the use-level to be able to classify different statements as either biased or not. Students may be asked to justify the classifications they make in cases where the appropriate classification is a matter of interpretation ("Is our school an example of a 'bureaucracy' or not?").

When students attain principles at the **use-level**, they learn to use them to build explanations, predictions or evaluations. Instead of learning simply to recognize or restate principles of physics, for instance, they would learn to use them to explain events in the physical environment, e.g., why a balcony in a newly constructed auditorium collapsed, or to predict what might happen in the future, e.g., whether next year's winter will be warmer than this year's. Principles that take the form of standards, e.g., "All citizens are entitled to due process of law," may be used to judge the rightness, worth, effectivenss or appropriateness of things.

When students attain procedures at the **use-level**, they are able to use them to solve problems, reach goals, and create products. Instead of learning about writing an essay, repairing a fender, cooking a turkey, or operating a micro-computer, for example, students learn to carry out these tasks.

Even more demanding than performance at the use-level is performance at the **refine**-level. At this level students are able to modify and extend content. That is, they are able to adapt and improve upon concepts, principles, or procedures that they have learned. For example, in a math class, a student might create a better approach to solving a particular type of problem. In social studies, a student might suggest amendments to the charter of a governmental agency. In fine arts, a student might refine a technique for using light and shadows in oil paintings. In the role of family member, a student might change his or her pattern of relating to siblings.

In general terms, one might say that students who are able to remember content are **knowledgeable.** Students who are able to use content are **skilled**, and students who are able to refine content are **adaptive**. Skill depends on knowledge, just as adaptiveness depends on skill. Similarly, students learn to perform at the remember-level before they are able to perform at the use-level, just as they achieve at the use-level before achieving at the refine-level.

Relation to Bloom's Taxonomy

This discussion of types of content and levels of performance to be attained stems from a model developed by David Merrill, a professor at the University of Southern California, and his associates (Merrill, 1983; Reigeluth, 1980). The model has been used successfully in the design of numerous instructional programs and accompanying assessment systems (Ellis & Wulfeck, 1982; Merrill et al., 1979).

However, the model is not as widely known as the system for classifying expected learning outcomes developed by Benjamin Bloom and his colleagues (Bloom, 1956; Bloom et al., 1971; Bloom et al., 1981). Bloom's "taxonomy," as the classification system is called, identifies six main categories of learning outcomes. The categories are viewed as cumulative in that behaviors at each level contain those of lower levels.

The first level in the taxonomy, termed **knowledge**, involves the process of receiving and remembering information, e.g., acquiring and restating definitions of concepts. Bloom's knowledge level is similar to the remember-level described in this handbook. One might say that being able to remember something and being knowledgeable about it are two sides of the same coin.

The other five levels of outcomes in Bloom's taxonomy deal with intellectual skills and abilities. Level 2, **comprehension**, involves skills like translating content from one form to another, e.g., converting a verbal statement into an equation, and drawing inferences and implications from information given, e.g., "the ability to extrapolate from data presented in a table" (Bloom et al., 1971, p. 476).

Level 3, **application**, refers to the ability to use "abstractions in particular and concrete situations," e.g., "The student can predict what will happen in a new situation by the use of appropriate principles and generalizations" (Bloom et al., 1981, p. 238).

Level 4, **analysis**, involves breaking down ideas and products into their component parts and determining the relations among the parts, e.g., "the student can use criteria (such as relevance, causation, and sequence) to discern pattern, order, or arrangement of material in a document" (Bloom et al., 1981, p. 252).

Level 5, **synthesis**, calls for "the putting together of elements and parts as to form a whole," e.g., "the ability to design a building according to given specifications" (Bloom, 1956, p. 171)."

Level 6, **evaluation**, is "the making of judgments about the value, for some purpose, of ideas, works, solutions, methods, materials, etc. . ." (Bloom, 1956, p. 185), e.g., "the ability to apply given criteria (based on internal standards) to the judgment of the work" (Bloom, 1956, p. 189)."

Bloom's taxonomy illuminates in a powerful way the range of cognitive learning outcomes that may be expected from instruction. However, there are two reasons an alternative model has been adapted for use in this handbook. One reason is that people not well versed in the subtleties of Bloom's taxonomy sometimes find it difficult to differentiate among the six levels. For example, the line between comprehension and analysis is at times difficult to discern.

Consider these two outcome statements:

1. "The student will be able to comprehend the significance of particular words in a poem in the light of the context of the poem" (Bloom et al., 1971, p. 412).
2. "The students can infer particular qualities or characteristics not directly stated from clues available in the document" (Bloom et al., 1981, p. 252).

According to Bloom and his colleagues, the first outcome statement illustrates comprehension. The second illustrates analysis. Yet the first appears to call for drawing inferences from context - a process very similar to that called for in statement 2.

Similarly, the distinction between application and evaluation is difficult to see in some cases. For example, a test item that assesses students' ability to determine the validity of a generalization, its relevance, and whether it supports or questions a trend, is classified in a text by Bloom and his colleagues (Bloom et al., 1981, p. 240) as a test for application. Later in the text, it is indicated that "the ability to distinguish between valid and invalid inferences, generalizations, arguments, judgments, and implications" is an evaluation skill (Bloom et al., 1981, p. 277). There often is a fine line between performance at one level of the taxonomy and performance at another.

A second reason for building upon the framework developed by Merrill as an alternative to Bloom's is that Merrill's framework brings into sharper focus the content underlying skills. The five levels of Bloom's taxonomy beyond the knowledge level all deal with "modes of operation and generalized techniques for dealing with material and problems" (Bloom, 1956). The contribution that knowledge makes to these general skills and abilities, however, is not addressed in the taxonomy.

In the framework presented in this handbook, skills are viewed in relation to knowledge. The skill of explanation, which would be classified as an analytic skill in Bloom's model, for example, is viewed as growing out of the acquisition and application of principles regarding cause and effect relations, just as skill in evaluation is seen as involving the application of different kinds of standards to various products, actions and conclusions. Explanation and evaluation may be "modes of operation," or "generalized techniques," but they cannot be taught effectively without clarity about the types of content which support them.

This is not to deny that some students are better thinkers or doers than others regardless of the specific knowledge they have acquired and applied. It simply is to recognize the strong contribution that knowledge makes to skillful and adaptive performance. Bloom's taxonomy does not focus attention on this connection.

The reader is encouraged, however, to consult the references on taxonomies of learning listed below, particularly the newest book by Bloom and his colleagues entitled Evaluation to Improve Learning (N.Y.: McGraw Hill, 1981). In this book, literally dozens of learning outcomes keyed to each level of the taxonomy are presented, as are corresponding test items. Particularly in the area of "synthesis," which involves creative production of ideas, art work, and other products - an area in which Merrill's framework is weak - the discussion and illustrations in Bloom's book are highly valuable.

REFERENCES

Chapter 2

Bloom, B.S. (Ed.) Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain. New York: McKay, 1956.

Bloom, B.S., Hastings, J.T. & Madaus, G.F. Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.

Bloom, B.S., Madaus, G.F. & Hastings, J.T. Evaluation to improve learning. New York: McGraw Hill, 1981.

Ellis, J.A. & Wulfeck, W.H. Handbook for testing in Navy schools (NPRDC Special Report 83-2). San Diego, Ca.: Navy Personnel Research and Development Center, 1982.

Merrill, M.D. Component display theory. In C.M. Reigeluth (Ed.), Instructional design theory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

Merrill, M.D., Reigeluth, C.M. & Faust, G.W. The instructional quality profile. In H.F. O'Neil, Jr. (Ed.) Procedures for instructional systems development. New York: Academic Press, 1979.

Reigeluth, C.M. The instructional quality profile: Training teachers to teach effectively. Educational Technology, 1980, March, 7-16.

RELATED RESOURCES

Doyle, W. Academic work. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1982.

Gagné, R.M. The conditions of learning, 3rd ed. New York: Holt, Rinehart and Winston, 1977.

Gagné, R.M. & Biggs, L.J. Principles of instructional design. New York: Holt, Rinehart and Winston, 1979.

Popham, W.J. & Baker, E.L. Systematic instruction. Englewood Cliffs, N.J.: Prentice Hall, 1970.

Posner, A.J. & Rudnitsky, A.N. Course design: A guide to curriculum development for teachers. New York: Longman, 1978.

CHAPTER 3

DIMENSIONS OF CONTEXT THAT INFLUENCE TEACHING AND TESTING

Introduction

Teaching and testing practices are influenced by more than desired learning outcomes. They also are influenced by the context in which these practices are carried out. Four aspects of context are discussed in this chapter:

1) students' learning backgrounds, abilities, attitudes and characteristics;
2) the instructional models a teacher has adopted;
3) the resources available for teaching and testing; and
4) the opportunities available for professional development.

Student Factors

Much of what a teacher does depends on the kinds of students with whom he or she is working. Students differ in terms of their past learning, the speed with which they are able to learn, their attitudes toward themselves as learners, their attitudes toward the content to be learned, their motivation to learn and a host of other characteristics that influence learning, e.g., their tolerance for ambiguity, dependence on authority, self-discipline, etc. Some of the implications of these differences among students for teaching and testing practices are outlined in Table 3-1 on the next page.

The importance of student backgrounds, abilities, attitudes and characteristics to teacher practice is addressed briefly in Chapter 6, on formulating expected learning outcomes, and in the chapters on using test information.

Instructional Models

The instructional model the teacher has adopted sets limits around the way in which teaching and testing are organized. Under a traditional model, instruction responds almost entirely to the content requirements of a text or curriculum. Evidence on student learning has a relatively small influence on instructional plans. Attention centers on what content has been covered rather than on what students know or are able to do in relation to this content. Under these conditions, tests are used primarily to assign grades to students. Test information plays a minimal part in instructional decision-making.

Efforts made to integrate teaching and testing within the framework of the traditional model of instructional management are clearly constrained. Definite limits are placed on a teacher's ability to adapt instruction to accommodate low performing and high performing students. A teacher cannot spend an extra two or three days providing corrective instruction or instruction intended to enrich and deepen a student's understanding of a topic, for example, if the press to cover a large number of topics is the primary concern.

In recent years, a number of instructional models have been developed that place less emphasis on the **extent** of content coverage and more on the **quality of students' learning**. As indicated earlier, these models include general approaches like "Mastery Learning" and "Active Teaching," and more specialized approaches, like "Jurisprudential Teaching" and "Learning Cycles." Mastery Learning and the Learning Cycles approach are discussed in the paragraphs that follow as illustrations of the influence an instructional model can have on teaching and testing patterns.

TABLE 3-1

The Effect of Individual Differences Among Students on Teaching and Testing Practices

Students With Limited Preparation in a Learning Area	**Students With Extensive Preparation in a Learning Area**
Instruction and assessment focus on rudimentary topics.	Instruction and assessment focus on more advanced topics.
Low Ability in a Learning Area	**High Ability in a Learning Area**
• The pace of instruction is slow.	• The pace of instruction is more rapid.
• Instructional presentations provide a large degree of structure to guide students.	• Instructional presentations may require students to make relatively long leaps of inference and discovery.
• Tests afford a high chance of success.	• Tests may be very demanding, stretching students to the limits of their skill and understanding.
• Tests, quizzes, and other assessment exercises are given frequently.	• Tests and related assessment devices may be given less frequently.
Low Interest/Motivation in a Learning Area	**High Interest/Motivation in a Learning Area**
• Special effort is made to tie content to the personal experience of students and to real-life situations.	• Less structure and effort is needed to demonstrate the meaningfulness of content to students.
• Students may be given extensive opportunity to demonstrate learning accomplishments in a form that reflects personal style and preference.	• Less personalized instruction and assessment is needed for students to learn.

SBB3

Mastery learning provides a general framework for planning and managing instruction. The model emphasizes the importance of giving each student sufficient opportunity to accomplish designated learning goals.

In mastery learning, the curriculum is clearly sequenced so that basic knowledge and skills are well developed before advanced content is introduced. Students are assessed prior to instruction to determine precisely at what point in the curriculum they should begin work. Instructional time is free to vary to accommodate the different rates at which students learn. Similarly, a variety of materials and strategies is used to accommodate differences in student learning styles. Tests are used routinely to monitor student progress toward attaining the learning outcomes identified in the curriculum, and to identify needs for instructional adjustment. Mastery tests are given upon completion of an instructional unit. These tests indicate whether a student is ready for new work, or whether instructional "recycling" is needed.* Tests thus play a key role in mastery learning, for they help a teacher decide where to begin instruction, when to move forward to a new topic or unit, when a topic or unit has to be taught over again, or alternative instructional procedures are needed. The effectiveness of mastery learning has been demonstrated in a number of studies (Block, 1974; Bloom, 1976, Ryan & Schmidt, 1979), including a study of a mastery learning approach to teaching high school chemistry (Swanson and Denton, 1976).

The **learning cycles** approach to instruction is an example of a more specialized instructional model. Whereas mastery learning is intended to apply across a wide variety of types and levels of learning goals, the learning cycles framework is designed to foster students' development of the specific kind of reasoning patterns that Piaget has described (although the model has much in common with general "inquiry" approaches to instruction).

At the secondary school level, the learning cycles approach has been most fully developed in the natural and physical sciences, through the work of Karplus and his associates at the University of California at Berkeley (Karplus, et al., 1980). When implemented in high school, the model is intended to foster students' development of "formal" patterns of reasoning. Formal reasoning, according to Piaget's conception, involves a number of key intellectual abilities, including:

---

* In subjects that are not tightly structured, like literature or international relations, the knowledge and skills dealt with in one unit may not be prerequisite for learning in later units. When mastery learning is applied in these subject areas, students often are permitted to advance to new material even if they do poorly on a unit test. In such cases, students generally are expected either to return later to the unit in which they had difficulty, or to try to achieve simultaneously the goals for a new unit and the goals yet to be achieved from the previous unit.

1. The ability to understand and apply abstract concepts, that is, those that refer to things that cannot be observed directly, but must be inferred, for example, the concept of "gene" in biology; the concept of "natural right" in political theory; and the concept of "ego" in psychology;
2. The ability to imagine all possible combinations of factors in making predictions, building explanations, or solving problems;
3. The ability to recognize and apply functional relationships, such as direct and inverse proportion;
4. The ability to separate the effects of several variables by varying only one at a time, for example, investigating whether temperature, moisture, or light intensity is most influential in determining the behavior of a beetle by designing experiments in which each of these factors but one is held constant (Karplus, et al., 1980); and
5. The ability to reflect on one's own thinking and recognize inconsistencies, bias, or other weaknesses in one's reasoning.

The learning cycles model suggests three phases of instruction that should be followed to foster students' development of formal thinking abilities: **exploration**, **concept introduction**, and **concept application**. During exploration, students are presented with the kind of problem with which they will be dealing throughout a particular unit of instruction. They investigate this problem, often in small groups, with minimal guidance from the teacher or expectation of specific accomplishment. The problems are intended to stimulate students' thinking about a particular learning area and to help make them aware of what they understand and don't understand in this area.

In the concept introduction phase, students are introduced to the new content that is needed to deal effectively with the kind of problem posed during exploration. In theory, the exploration phase stimulates students' interest in this new content. The teacher's role during concept introduction is to explain the content and assure that students can relate it to the initial exploratory activity.

In the concept application phase, students use the new concepts or principles in a range of situations. This helps them to generalize their understanding beyond the examples provided during exploration and concept introduction.

With respect to assessment, the learning cycles model suggests the importance of using problem solving tasks and essays that call for formal reasoning on students' part and which require them to explain or justify their conclusions or answers. In the materials prepared by Karplus, for example, many excellent examples of such procedures are furnished, accompanied by guidelines for evaluating students' responses. The model also suggests that informal assessment, e.g., observing small group investigations or probing students' responses during discussion, is most appropriate during the exploration and concept introduction phases, and that more formal assessment of students' learning is most appropriate when the application phase is completed.

Differences among a traditional model of instruction, mastery learning, and the learning cycles approach are summarized in Table 3-2.

Resources Available to the Teacher

High quality resources for teaching and testing also are important aspects of context.

Instructional Resources

The assumption is made in the handbook that teachers will select instructional material, e.g., texts, films and computer-assisted programs, that are directly related to established learning goals. Unfortunately, in some settings teachers lack the materials needed to foster students' attainment of key goals. Though teachers can and do create some of the learning materials they need, they stand a better chance of meeting instructional objectives when appropriate materials are close at hand.

The assumption also is made in the handbook that teachers will make an effort to accommodate students whose performance on tests is below expectations or consistently above them. However, teachers who lack a wide range of instructional materials can go only so far in adjusting instruction in view of individual differences in test performance. In a major research study of teaching practices in American high schools, large numbers of teachers reported that the main reason they seldom focused instruction on the needs of individual students was the lack of instructional resources (Weiss, 1978). Varied instructional resources assist teachers in responding flexibly to test results.

Test-Related Resources

Teachers are more likely to use tests in planning and decision-making if they have access to test-related resources. Test-related resources include tests or test item pools that are directly linked to the curriculum with which teachers are working. (These could include essay topics, problems, and open-ended questions that have proved useful to teachers in the past, as well as the more common response-selection type of test item.) Test item banks permit teachers to generate tests that serve a variety of purposes, for example, pretesting, progress-checking, posttesting, end-of-year testing. In some districts, teachers are provided with released time to work with colleagues in developing and pilot testing test items to be included in an item bank. Commercial

Table 3 - 2

Different Instructional Models and Their Implications for Teaching and Testing

ASPECTS OF TEACHING & TESTING INFLUECNED BY MODELS		MODELS	
	Traditional Model	Mastery Learning	Learning Cycles
NATURE AND FORMULATION OF LEARNING GOALS FOR STUDENTS	Topics or themes to be covered typically are identified, but learning outcomes to be achieved generally are not.	Specific goals and objectives for courses and units are established. Relationships among different goals and objectives are analyzed carefully.	Relatively general goals are established that focus on broad patterns of reasoning and key concepts and principles. Specific objectives may or may not be formulated.
APPROACHES TO INSTRUCTION	Instruction tends to be expository in nature, consisting largely of lectures and question-answer sessions, and tends to be anchored tightly to textbooks.	Instructional strategies are matched to the specific objectives to be mastered. If evidence suggests these strategies are less effective than desired, new strategies are used.	Instruction follows three phases: exploration, concept introduction, and concept application. Extensive use is made of small group investigations and discussions.
APPROACHES TO ASSESSMENT	Assessment typically involves brief responses on the part of students, e.g., recall of facts, terms or rules. When assessment does focus on higher-order learning, students often are unprepared because such learning has not been fostered during instruction.	Assessment formats are matched to the type of objective to be assessed.	Assessment formats vary, but particular use is made of problem solving tasks and essays which require students to justify their conclusions and answers.
USES MADE OF ASSESSMENT INFORMATION	Assessment information is used primarily for grading.	Assessment is used for a variety of purposes: -to determine students' "entry level" knowledge and skills as a guide for student grouping and instructional planning; -to provide feedback to students on their learning progress and to determine if and when "corrective" instruction is needed; -to assess students' mastery at the end of instruction; -to assess the effects of corrective instruction on students' learning; -to assign grades.	Although various uses of assessment are made, special emphasis is placed on using assessment responses as a means for stimulating discussion among students as to the most appropriate or effective strategy for dealing with a particular kind of problem or issue.

publishers and nonprofit evaluation associations also have created item banks.

Insofar as teachers wish to assess students' aptitudes, standardized, norm-referenced tests also constitute important resources. Machines that score tests and analyze test results also are helpful, particularly the small, desk-top variety, as are aides that assist in the scoring and analysis process. Though most high school teachers probably will rely on their own tests to monitor and evaluate student learning, and probably will do most of the scoring and analysis of tests themselves, supplementary resources are likely to facilitate teachers' work in these respects (Hambleton, Anderson, & Murray, 1983).

Testing resources of a more specialized nature have a large potential for testing and teaching. For example, self-scoring and self-diagnosis procedures have major implications for classroom practice. These are procedures in which students score tests themselves, and are provided with an interpretive booklet that offers explanations of why each keyed answer is considered preferable to the alternatives. The College Board's Career Skills Assessment Program, for example, features tests of this kind. Answer sheets accompanying the program are designed so that as soon as the student has completed the test the layered answer sheets can be separated. Students can immediately score their own answers on one part; the other part can be machine scored. The accompanying test booklet gives a rationale for each answer and presents illustrations of how the skills assessed on the text can be applied in practice.

Specialized testing resources often are built into larger instructional systems. In these systems, tests may be so closely connected to learning that they represent a form of teaching. For example, in some computer-assisted programs students are asked to respond to simulations of real-life problems. For each response students make on the path toward reaching a solution, the computer provides feedback on the consequences of their choice. The computer also poses different questions to students depending on their responses to earlier questions. This approach to testing and teaching has been used for years in computer-assisted training programs for pilots. Simulators for other learning problems are being developed for use with high school students (Krumboltz, 1982).

Opportunities for Professional Development

In some school districts teachers have many opportunities to improve their teaching and testing practices, and to influence aspects of the context in which these practies are carried out. These opportunities might include released time or summer stipends for curriculum development and test construction, or the preparation of fully integrated instructional units. Further opportunities and incentives might be provided to work with fellow teachers and administrators on program evaluation teams, or in policy making groups concerned with competency testing, school or department grading policies and the like. In addition, staff development programs might be arranged through which teachers can extend their skills in various elements of instruction and assessment, or in using one or more of the models of effective instruction referred to earlier. The type and extent of a district's support for teachers' professional development has a far-reaching influence on the teaching-testing process.

REFERENCES

Chapter 3

Block, J. H. (Ed.) Schools, society, and mastery learning. New York: Holt, Rinehart, and Winston, 1974.

Bloom, B. S. Human characteristics and school learning. New York. McGraw Hill, 1976.

Good, T.L., & Grouws, D.A. The Missouri mathematics effectiveness project: An experimental study in fourth-grade classrooms. Journal of Educational Psychology, 1979, 71 (3), 355-362.

Hambleton, R.K., Anderson, E.G., & Murray, L.N. Applications of microcomputers to classroom testing. In W.E. Hathaway (Ed.) Testing in the schools. San Francisco, Ca.: Jossey-Bass, 1983.

Karplus, P. & Others. Science teaching and the development of reasoning. Berkeley, CA: Lawrence Hall of Science, University of California, 1980.

Krumboltz, J. D. Tests and guidance: What students need. In W. B. Schrader (Ed.), Measurement, guidance, and program improvement. San Francisco, Ca.: Jossey-Bass, 1982.

Ryan, D.W., & Schmidt, M. Mastery learning: Theory, research, and implementation. Ontario, Canada: Ministry of Education, 1979.

Swanson, P., & Denton, J. A comparison of remediation systems affecting achievement and retention in mastery learning, 1976 (ERIC Document Reproduction Service No. Ed 131 037).

Weiss, I.R. National survey of science, mathematics, and social studies education. Washington, D.C.: National Science Foundation, 1978.

RELATED RESOURCES

Anderson, L.W. Assessing affective characteristics in the schools. Boston: Allyn and Bacon, 1981.

Arlin, M., & Webster, J. Time costs of mastery learning. Journal of Educational Psychology, 1983, 75 (2), 187-195.

Fenstermacher, G.D. and Goodlad, J.I. (Eds.) Individual differences and the common curriculum. (Eighty-second yearbook of the National Society for the Study of Education, Part I). Chicago: University of Chicago Press, 1983.

Klausmeier, H.J., Lipham, J.M., & Daresh, J.C. The renewal and improvement of secondary education: Concepts and practices. Lanham, M.D.: University Press of America, 1983.

Messick, S. & Associates. Individuality in learning. San Francisco, Ca: Jossey-Bass, 1976.

Moos, R.H. (Ed.) *Evaluating educational environments*. San Francisco, Ca.: Jossey-Bass, 1979.

Student learning styles: Diagnosing and prescribing programs, Reston, Va.: National Association of Secondary School Principals, 1979.

Talmage, H. (Ed.) *Systems of individualized education*. Berkeley, Ca.: McCutchan, 1975.

SBB3

CHAPTER 4

PURPOSES SERVED BY TESTS

Introduction

In the preceding chapter on the context in which teachers work, a number of models of instructional effectiveness were mentioned. From these models, a common set of purposes for classroom testing can be derived. The intent of this chapter is to draw out in explicit form these purposes.* They include:

1. To assess the learning accomplishments of students prior to instruction or placement;
2. To assess the progress students are making toward particular learning goals during a course or school year;
3. To identify the causes of poor performance by students; and
4. To determine the extent to which long-term expected learning outcomes have been accomplished.

Tests administered on a district-wide basis serve somewhat different purposes than the four that have been identified. Several purposes related to district managed testing are discussed briefly in the last section of the chapter.

---

* Questionnaires, surveys, interviews, and other devices that might be used to assess student attitudes and learning characteristics are not considered tests in the context of the handbook, and consequently are not addressed in this chapter.

Purposes Served By Classroom Tests

Assessing Students' Past Learning

Tests of students' past learning may provide useful information in establishing appropriate learning expectations and in designing appropriate instruction. All the major models of instructional effectiveness indicate the need to assess students' learning accomplishments prior to a course or unit as a basis for instructional planning.

Various aspects of students' past learning may be assessed. Teachers may, for example, design tests to check whether students have the knowledge and skills needed to do the work called for in a particular unit or course. In some districts, placement tests may be administered to assure that students entering a course are equipped to succeed in it, but even in these cases teachers may wish to confirm results from placement tests or to develop tests focusing on prerequisites for a particular unit. For example, before starting a unit on the Reconstruction period in American history, a teacher may wish to assess whether students have basic information about the Civil War and its background.

Teachers may also design tests to assess what students know or can do in relation to topics to be covered. For example, a health course may focus on many topics about which students have considerable knowledge, like nutrition, common medical problems, the negative effects of cigarettes, etc. A pretest in cases like this helps the teacher identify what new topics or subtopics need to be addressed. In individualized instructional programs, pretests commonly are used to assure that each student is assigned instructional material at an appropriate level of complexity or difficulty.

Assessing Students' Learning Progress

A vitally important purpose of classroom tests is to provide information on the progress students are making toward attaining desired learning outcomes.

Information from progress tests is clearly beneficial to students. It confirms and reinforces what and how well they have learned, or identifies learning gaps that need to be filled. There is a sizable body of research indicating that the frequency of quizzes and other forms of testing, when accompanied by detailed feedback to students, is related to the final performance level of students (cited in Bloom et al., 1981, p. 170). Different types of students and different types of learning outcomes may call for different levels or types of testing. In one form or another, however, information on learning progress being made is important to all students.

Progress testing also is important for teachers. Information from progress tests may suggest the need to review material with students, to change instructional procedures or sets of materials, or to assign students more basic or more advanced work. In classrooms where teaching and testing are integrated, instructional decisions hinge to a large degree on the assessment of learning progress.

Identifying Causes of Poor Performance

Tests may be used to investigate the reasons why a student or group of students is performing poorly in a learning area. Teachers may analyze the pattern of errors that students make on a test to help pinpoint underlying misconceptions. Teachers also may design special tests to investigate the causes of students' poor performance. For example, a physics teacher may give a test on the basic skills of mathematics to determine whether students' poor performance on physics problems results from a lack of prerequisite math skills.

When students repeatedly experience failure on learning tasks, and teachers are unable to identify the sources of the difficulty, the students generally are referred to counselors and other specialists for more extensive and refined diagnostic assessments. Standardized diagnostic instruments usually are employed in these cases. In this handbook, however, attention centers on the less formal diagnostic techniques teachers create themselves to detect students' underlying misconceptions and error patterns.

Assessing the Attainment of Long-Term Learning Outcomes

Another purpose served by tests is to assess what and how well students have learned at the end of a course or school year. Tests that focus on outcomes expected at the end of a long period of instruction often are referred to as "summative" measures of achievement.

Results from summative evaluations generally weigh heavily in a students' grade. Since the consequences of scoring high or low on summative tests are large, these tests must be constructed and administered with particular care.

Purposes Served By District-Managed Tests

The focus of this handbook is on classroom testing. However, tests given by a school district influence the context in which instruction is carried out and may also provide information that is useful for instructional planning.

Tests administered by a school district generally serve one of four purposes:

1. To determine the most appropriate **placement** of a student, either in a program or course. For example, results from placement tests may help teachers determine which learning group a particular student should be placed in;

2. To **certify** that students have the knowledge and skills needed to succeed in a particular role or context. For example, district managed competency tests are often used to establish that a student has the basic skills needed to function effectively in life roles;

3. To **evaluate program effectiveness.** For example, a common essay exam given to all tenth grade students in general writing classes may provide useful information on the effectiveness of a school's writing program; and

4. To **select** students who are most likely to succeed in a particular environment, like a four-year college or the armed forces. Whereas placement tests determine where a student is to be assigned, selection tests determine whether students' qualify, or will be chosen, for one or more opportunities. Selection tests usually are designed and used by agencies external to the school, but they are widely accessible through school districts and sometimes even influence a district's curriculum.

REFERENCES

Chapter 4

Bloom, B. S., Madaus, G. F. & Hastings, J. T. *Evaluation to improve learning*. New York: McGraw Hill, 1981.

RELATED RESOURCES

Baker, E.L. *Issues in achievement testing* (CSE Resource Paper No. 3). Los Angeles, Ca.: Center for the Study of Evaluation, University of California, Los Angeles, 1982.

Buros, O.K. Fifty years in testing: Some reminiscences, criticisms and suggestions. *Educational Researcher*, 1977, *6*, 9-15.

Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. *Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs?* (CSE Report No. 1941). Los Angeles, Ca.: Center for the Study of Evaluation, University of California, 1982.

Competency tests and graduation requirements. Reston, Va.: National Association of Secondary School Principals, 1979.

Hathaway, E.W. (Ed.) *Testing in the schools.* San Francisco, Ca.: Jossey-Bass, 1983.

Jaeger, R.J. and Tittle, C. (Eds.). *Minimum competency achievement testing*. Berkeley, Ca.: McCutchan, 1979.

Phi Delta Kappan, 1981, May, Special section on standardized testing. 623-636.

Popham, W.J. (Ed.). *Evaluation in education: current applications.* Berkeley, Ca.: McCutchan, 1974.

Tyler, R.W. and Wolf, R.M. (Eds.). *Crucial issues in testing*. Berkeley, Ca.: McCutchan, 1974.

SBB4

CHAPTER 5

TYPES OF TEST ITEMS AND ASSESSMENT PROCEDURES

Introduction

This chapter provides an overview of the most commonly used types of test items and assessssment procedures. Three broad approaches to assessment are addressed: **Objective** test items; **essay questions** and other tasks that require students to create products; and **observation of performance.** Objective items are those that can be scored with little or no subjective judgment, e.g., multiple choice and fill-in-the-blank items. Essay questions and other product-related forms of assessment require teachers to make judgments about quality when evaluating students' work. Students may be asked to produce a wide range of products for assessment purposes, including sculptures, paintings, objects from wood, mechanical drawings, laboratory reports, etc., which are then evaluated for their quality.

Observation of performance involves the assessment of student behaviors as they occur, for example, assessing a student's skill in dribbling a basketball, delivering a speech, driving a car, using a microscope, or leading a discussion. While all tests require students to perform a task, tests calling for observation of performance focus on performance "in process," that is, performance as it takes place.

It should be noted that the handbook focuses on observation as an approach to achievement testing. When observations are used for this purpose, they are conducted in structured situations and guided by rules for recording and evaluating student's responses. This is not to suggest that informal or unstructured observations lack importance, but only that they are not of primary concern in the context of the handbook.

Knowledge of various approaches to testing is essential for preparing tests that are appropriate to the different types and levels of learning outcomes desired from instruction, and that are consistent with the time and resources available for test administration and scoring.

Objective Tests Items

Objective test items can be scored with minimal or no subjective judgement. They are designed according to either a **response-selection** or a **response-completion** format.

Response-Selection Format

A response-selection format requires students to choose an answer from a set of options. "Multiple choice," "alternative response," and "matching" items exemplify the response-selection format. A description of each of these item forms, along with illustrative test items, are presented in Table 5-1 on the following page.

Response-Completion Format

A response-completion format requires students to construct a response rather than choose one. "Fill-in-the-blank" and "short answers or exercises" exemplify the response-completion format.* Although teachers often must use judgment to determine which responses to these items are acceptable and which are not, the degree of judgment required is small (assuming the items have been carefully constructed). For this reason, fill-in-the-blank and short answer items are classified as objective items in the handbook. A description of each of these item forms, along with illustrative test items, are presented in Table 5-2.

Essays and Other Product-Related Tasks

Essays and other product-related tasks are particularly well-suited for assessing student's ability to integrate and apply content they have been studying. Essays are frequently used in courses in the humanities and social sciences. Product-related tasks, in which students produce concrete pieces of work, e.g., a mechanical drawing, a musical composition, an article of clothing, are used in a variety of subject areas. An example of an essay test and a product-oriented test are presented on page 38.

---

* Items referred to as "response completion items" in this handbook are often treated in the literature on testing under the more general category "supply type items." Supply type items include essays. In this handbook, however, an essay is treated as a separate type of measure because of its distinctive properties. The term "response - completion items" is intended to refer only to objective items.

TABLE [illegible]-1

An Overview of Response - Selection Test Items

Type of Item	Description	Illustrative Items	
Multiple choice	This type of test item asks students to select the correct or best response to an incomplete statement, question or problem. Generally, four to five response options are provided. Only one option is correct; the others are "foils" or "distractors" that represent plausible but incorrect answers.	Which one of the following statements illustrates alliteration? A. The bees buzzed. B. What a foolish fancy he had. C. Their emotions peaked. D. His legs were like stumps.	The Pucker-Up Lemonade Stand started the day with all its containers full of lemonade. It had 2 gallon containers, 3 quart containers, and 2 pint containers. What is the maximum number of oup servings that could be served in a day? A. 96 B. 48 C. 32 D. 22 E. 7
Alternative response	This kind of test presents students with a choice of two response options. The options generally are direct opposites, e.g., true/false, yes/no, correct/incorrect, valid/invalid. But they also can represent different but closely related concepts, for example, metaphor/simile, affectionate/sentimental, happy/esctatic.	Hydrogen gas is highly combustible. True False	Brahms was the greatest composer of the 19th century. Fact Opinion
Matching	A matching test is composed of two lists, e.g., of names, places, events, characteristics, etc. Students must match the terms of one list to the terms of the other. One list generally has more terms than the other so correct responses cannot be chosen by a process of elimination.	Write the letter of the sport in column 2 that matches each athlete in column 1. Column 1 __1. Babe Ruth __2. Billy Jean King __3. Mark Spitz __4. Joe Namath	Column 2 A. swimming B. football C. tennis D. skiing E. baseball F. golf

TABLE 5-2

An Overview of Response - Completion Test Items

Type of Item	Description	Illustrative Items
Fill-in-the-blank	A fill-in-the-blank item, sometimes called a completion item, consists of a statement in which a key element, e.g., a term, concept, number, place, etc., has been left out. Students are asked to complete the statement by supplying the key element.	In 1982, Britain waged war against Argentina in a dispute over the (Falklands or Malvinas) Islands.
Short Answer or Exercise	This type of item asks students to supply a brief response (a word, a phrase, a couple of sentences) to a direct question, to carry out a specific activity, or to solve a particular problem that has only one correct solution.	. Sally is taller than Sue. Sue is taller than Jane. Who is the shortest of the three girls? . State in your own words the central constitutional issue involved in the 1966 Supreme Court case of Miranda vs. Arizona. . Draw a circle with a diameter of 2 inches. . In building a bridge, a 493 foot cable is needed. Since you can only buy cable to the nearest 10 feet, how long a cable must be purchased?

Illustrative Essay Test: European History Class

Write a two to three page essay (no more than 750 words) on the reasons underlying the collapse of the Weimar Republic. Be sure to discuss political, economic and cultural factors that contributed to the breakdown. You will have the entire period to complete the essay.

The essay is worth 9 points. Three (3) points will be awarded for the organization of the essay (its clarity and coherence); three (3) points will be awarded for the quality of evidence used in the essay (its relevance and accuracy); and three (3) points will be awarded for the depth of insight reflected in the essay, i.e., the degree to which fundamental, as opposed to superficial, causes of the Republic's collapse are addressed.

Illustrative Product-Related Test: Mechanical Drawing Class

Draft a blueprint for a tool shed that meets the following specifications:

- The shed should be able to withstand variations in temperature from -10° Farenheit to 100° Farenheit;
- It should be large enough to house three wheelbarrows, a sit-down lawnmower, and a dozen rakes, tools and other instruments; and
- The materials for the shed should cost no more than $700.

Essay and other product related assessment procedures vary according to the degree of freedom provided to students in composing a response. The essay question about the collapse of the Weimer Republic presented above, for example, provides some structure to guide students' responses, i.e., the broad categories of causes students are to discuss are specified, as does the illustrative product-related test, but neither sets tight response boundaries.

It is possible to develop tests of this kind that permit a great deal of freedom of response, e.g., "Think of a person whom you admire. Write an essay to describe the person and explain why you admire him or her" (Priestley, 1982, p. 196). Tests that allow such extensive response freedom typically are used to measure general essay writing or product-design skills. Tests that set tight response boundaries are generally used to measure achievement in a focused content area.

Tests that set tight response boundaries usually are easier to score than those permitting greater response freedom. Scoring keys that indicate the characteristics to be included in an essay or product, and the point values assigned to each, can be developed more readily when the nature of a desired response has been specified in some detail in the test item. This advantage, however, needs to be weighed against its potential disadvantage. If too little freedom is provided, an essay or other form of product-related test may be reduced to a series of short answers or exercises which do not require students to integrate or apply material (Brown, 1981, p. 66).

Scoring essay and similar tests, even those that carefully structure students' responses, necessarily requires more subjective judgment than scoring short answers or exercises. Procedures have been established, however, to reduce subjectivity in scoring. These involve the use of scoring keys and two commonly used derivations of scoring keys: **checklists** and **rating scales.** Checklists are useful for assessing whether a particular characteristic is present or absent in an essay or other product. Rating scales are useful for assessing the degree to which a characteristic is present or absent. An illustrative checklist is presented in Table 5-3 below. An illustrative rating scale is presented in Table 5-4 on the following page.

TABLE 5-3
An Illustrative Checklist for
Assessing a Product in Metal Shop

Spot-Welded Aluminum Case

Product Characteristics	Yes	No
1. Dimensions of the case match specifications	____	____
2. Six spot welds are visible	____	____
3. Edges have been filed	____	____
4. Metal has been cleaned	____	____
5. All corners are right angles	____	____

(Priestley, 1982, p. 136)

Observation of Student Performance

Wher[illegible] essays and similar forms of assessment focus on concrete products that result from student performance, observation, as a method of testing, involves the assessment of performance "in process," e.g., observing and evaluating how well students execute a double play in baseball or express their ideas in a discussion.

TABLE 5-4

An Illustrative Rating Scale To Evaluate Student Themes

Directions: Below are two sets of scales for judging themes. The first lists characteristics of content, the second lists those of form and style. Check the position in each scale that best represents your opinion of that characteristic of the theme. Add comments as desired.

Content

1. Quantity and quality of topic investigation

1	2	3	4	5
Very limited investigation; little or no material related to topic.		Fair amount of investigation; some material not well adapted to topic.		Extensive investigation; good selection of essential material on chosen topic.

Comments:

Form and style

1. Word Usage

1	2	3	4	5
Poor usage of words; few used correctly.		Fair usage of words; some not used correctly.		Good usage of words; practically all words used correctly.

Comments:

(adapted from Ahmann & Glock, 1967, p. 230)

SBB5

Observations are particularly useful for assessing students' ability to adjust their behavior in response to changing cues, conditions, issues, or events. Interpersonal communication represents a prime context for observing students' ability to respond sensitively and effectively to situational changes, since such responsiveness is the essence of two-way communication. Athletic competition and the use of complex machines, e.g., computers, automobiles, and life-sustaining medical equipment, also provide many instances for observing students' adaptive behavior.

Two factors make observations of student performance especially challenging. First, unlike products which are durable and can be evaluated at the teacher's convenience, performance is time and place bound. If a teacher wants to observe the degree to which a student empathizes with fellow students about school-related problems, the teacher must be present at problem-sharing situations where empathy is called for, or he/she must create a simulated situation. In either case, the teacher must be on hand to make the observations at a particular time.

The second factor that makes observations complex is the potential for observation to be obtrusive. The presence of the teacher in many situations disrupts the "naturalness" of the situation and promotes an artificial response on the students' part, as would be the case, for instance, in the example cited above about student empathy.

It is sometimes feasible for the teacher to observe student performance without cuing the students that they are being observed. A coach could watch a tennis match from some distance, for example, and make notes on a student's tennis skills. A teacher could unobtrusively keep a special eye on one or two students as they worked in small groups to observe their interpersonal communication skills. However, when observations are conducted as "tests" on which students will receive scores, it does not appear ethical to make observations without informing students that they are being observed.

More will be said about performance evaluations in Chapters 9 and 10. Suffice it to note at this point that the same type of instruments used to assess products, i.e., checklists and rating scales, are used to guide and score observations of student performance. In addition, frequency counts of specific student behaviors sometimes are used in performance evaluation. An illustrative rating scale for assessing students' discussion skills is presented in Table 5-5. An illustrative procedure for counting frequencies of behaviors displayed in returning volleys in tennis is shown in Table 5-6.

TABLE 5-5

Illustrative Rating Scales for Evaluating Two Discussion Skills

When, during a small group dicussion of a public issue, a student is given the opportunity to:

1. ask a question, he or she	Asks an obviously irrelevant question	Asks a question that is somewhat related to the topic under discussion	Asks a question that is directly related to the topic under discussion
	0	1	2
2. express disagreement with a speaker, he or she	Engages in a personal attack or ridicules the speaker	Refrains from from personal attacks, but does not clearly identify the reasons for the disagreement	Refrains from personal attack and clearly identifies the reasons for the disagreement
	0	1	2

SBB5

TABLE 5-6

Evaluating Volley Returns in Tennis

Directions: Below are criteria for evaluating volley returns. Whenever during a set a student returns a volley, or attempts a return, check the evaluative criteria below that apply. Circle the total number of volleys observed.

Number of volleys in the set	1 15 20	2 16 30	3 17 31	4 18 32	5 19 33	6 20 34	7 21 35	8 22 36	9 23 37	[✓ 10] 24 38	11 25 39	12 26 40	13 27 41	14 28 42	**Percentage of volleys in which a criterion is met**
Racket in FRONT								𝍸	///						80
Racket head HIGH								𝍸							50
Knees BENT, but head UP								𝍸	/						60
LEAN or STEP into shot								𝍸	///						80
Keep racket head moving FORWARD								𝍸	//						70
Watch the ball CONTACT the racket face								///							30
NO FOLLOW-THROUGH								𝍸							50
BE AGGRESSIVE								𝍸	///						80

(adapted from Hopkins & Antes, 1979, p. 173)

SBB5

REFERENCES

Chapter 5

Ahmann, J. S., & Glock, M. D. Evaluating pupil growth. Boston: Allyn and Bacon, 1967.

Brown, F. G. Measuring classroom achievement. New York: Holt, Rinehart, and Winston, 1981.

Hopkins, C. D., & Antes, R. L. Classroom testing: Construction. Itasca, Ill.: F. E. Peacock, 1979.

Priestley, M. Performance assessment in education and training: Alternative techniques. Englewood Cliffs, N.J.: Educational Technology Publications, 1982.

RELATED RESOURCES

Bloom, B.S., Madaus, G.F. and Hastings, J.T. Evaluation to improve learning. New York: McGraw Hill, 1981.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Gronlund, N.E. Measurement and evaluation in teaching. (4th ed.). New York: Macmillan, 1981.

Mehrens, W. A. & Lehmann, I. J. Measurement and evaluation in education and psychology, 3rd ed. New York: Holt, Rinehart and Winston, 1984.

TenBrink, T.D. Evaluation: A practical guide for teachers. New York: McGraw-Hill, 1974.

SBB5

PART II

APPLICATIONS

CHAPTER 6

FORMULATING EXPECTED LEARNING OUTCOMES

Introduction

In the first part of this handbook, the importance of expected learning outcomes for both instruction and assessment was noted. Expected learning outcomes influence the selection of instructional strategies and materials as well as the design of test items and assessment procedures.

Issues to consider when formulating expected learning outcomes are discussed in this chapter, and illustrative formats for writing outcome statements are presented. The chapter is organized around five aspects of formulating outcome statements: (1) determining the level of generality at which outcome statements are to be written; (2) deciding how much information to include in an outcome statement; (3) determining the relative emphasis to place on outcomes for purposes of instruction and assessment; (4) using tests to adapt expected outcomes to students' learning backgrounds and abilities; and (5) preparing outcome statements for students' use. It is emphasized in the chapter that statements of expected learning outcomes may take different forms, depending on the context in which they are developed and used.

Determining the Level(s) of Generality at which Outcome Statements are to be Written

Outcome statements can be written at different levels of generality. Highly general statements may be written in reference to broad segments of instruction, e.g., a program, a course, or a large unit. (General outcome statements are typically referred to as **goals**.) More specific outcome statements may be written in reference to narrower segments of instruction, e.g., a short unit or an individual lesson (specific outcome statements are typically referred to as **objectives**).

Here is an example of a learning goal and set of supporting objectives:

Goal

Students will be able to summarize concisely and accurately the content of nonfiction reading selections.

Objectives

- Given a paragraph containing a series of related items (e.g., Tokyo, New York, Mexico City, London, Bombay), students will be able to substitute for the series a superordinate term or phrase (large cities).
- Students will be able to distinguish between descriptive details and main ideas in passages from widely used science and social studies texts.
- Given an expository essay, students will be able to identify: (a) the thesis statement of the essay, and (b) the topic sentence of each paragraph in the essay.
- Students will be able to create a title for a news article that is consistent with the information and ideas presented in the article.

The above illustration shows the distinction between a learning goal and objectives related to it. However, knowing that an intended outcome is more general than another is one thing. Knowing exactly how general or specific an outcome statement should be is another. For example, the goal in the following example is clearly more general than the objectives intended to support it. However, the question remains whether the objectives are specific enough to serve as effective guides to instruction and assessment.

Goal

The accelerated learning group will use appropriate and effective processes to reach solutions to non-routine problems in math.

Objectives

- Members of the accelerated learning group will be able to:
  1. list the knowns and unknowns in a problem;
  2. distinguish between relevant and irrelevant information in a problem statement;
  3. infer additional information from what is given;
  4. develop an organized method for solving a problem; and
  5. change an approach to solving a problem if the approach being used proves unproductive.

For some teachers and some students, the processes indicated in this objective might be sufficiently clear and specific to guide teaching, learning, and assessment. However, these processes are relatively broad and abstract in nature. Undoubtedly, not all math teachers would be prepared to derive instructional strategies or assessment procedures from these objectives.

The appropriate level of generality of an outcome statement seems to depend on the nature of the learning area in which students are working, on the knowledge and skills students bring to a learning situation, and on a teacher's professional training. Hard and fast rules about the generality/specificity of outcome statements do not appear defensible.

Deciding How Much Information to Include in an Outcome Statement

Once the decision is made about the appropriate level of generality of an outcome statement, further thought must be given about the specific elements to include in the statement. Texts in instructional design commonly recommend that learning objectives include information about the **content** to be learned (facts, concepts, principles, and procedures); the **level of performance** to be attained (remembering content; using it; or refining it); the **conditions** under which performance is to occur or that set limits around it; and the **standard** of acceptable performance for the objective (the degree of proficiency, accuracy, or speed that must be reached to indicate that an outcome has been attained, and the number of students that must reach this level to establish that instruction has been successful).*

Here is an example of an objective that includes information on all four of these elements:

> Eighty percent of the students in the class will be able to identify in nine of ten passages from a contemporary novel the use of one or more of the following literary techniques: personification, hyperbole, paradox, irony, or satire.

In this example, the content consists of literary **concepts**, e.g., personification, hyperbole, etc. Students are expected to **use** the concepts, rather than merely remember them. The **condition** structuring performance is the information presented in the ten passages. The **standard** is that "eighty percent" will be able to identify in "nine of the ten passages" the author's use of a particular technique.

In this handbook the position is taken that while the content students are to learn and the level of performance they are to attain (remember, use, or refine) must be made clear in outcome statements, performance conditions and evaluative standards do not necessarily have to be spelled out in these statements. It is recognized that outcome statements generally are written before one has developed specific instructional procedures or specific test items or other assessment devices. Therefore, one may not know at the time outcome statements are being prepared exactly what conditions will govern performance on an assessment task, or what scores students will be expected to attain to indicate that an outcome has been achieved.

Consider, for example, the following objective: Students will be able to design a valid survey to assess the opinions of members of the community on selected public issues. Neither performance conditions nor evaluative standards are spelled out in this objective, yet the objective may serve effectively as a basis for designing instruction and assessment. The objective could be extended to include information on performance conditions, e.g., "After discussing the procedures for constructing survey instruments presented in the textbook, and applying these procedures in practice exercises, students will, without the aid of a text or other resource . . . " But such a description seems more appropriately placed in lesson plans or in directions for an assignment or assessment exercise than in a statement intended to convey the nature of the outcome to be achieved.

---

* In the literature on testing, references sometimes are made to "amplified" or "expanded" objectives. These objectives provide even more specific information on outcomes to be achieved than the objectives described and illustrated in most textbooks. For more information on amplified objectives, see the book by Baker and Popham and the chapters by Hambleton and Popham listed in the "Related Resources" at the end of this chapter.

Similarly, one could add to the illustrative outcome statement above information on the standards to be used in evaluating students' work, e.g., "A four-point rating scale, adapted from one presented in the students' text, will be used to judge the quality of students' surveys. A rating of 3 will be considered acceptable." However, in order to specify evaluative standards in this case a teacher would have to have developed or chosen a particular evaluation instrument at the time the objective was written. The objective thus would serve as a description of an assessment tool that already had been developed rather than as a guide to developing such a tool.

If the point of writing objectives is to establish direction for assessment, it seems unreasonable to expect objectives to describe what assessment procedures look like when they are completed. Teachers should not feel obligated to build into outcome statements descriptions of performance conditions or evaluation standards unless it is practical and meaningful to do so. This may be the case for highly specific objectives for which performance conditions and standards are readily apparent. In most cases, however, information on content and level of performance will provide a sufficient basis for planning instruction and preparing assessment instruments.

Determining the Emphasis to Place on Expected Outcomes for the Purpose of Instruction and Assessment

Teachers may wish to place more emphasis on some outcomes than others in instruction and assessment. An outcome may receive particular emphasis if it is considered essential (a) to subsequent learning, e.g., basic reading skill is essential in learning to analyze literature, or (b) in the successful performance of a life-role, e.g., knowledge of the first amendment is essential in carrying out the responsibilities of a citizen. An outcome also may be viewed as needing special emphasis if it is particularly complex or difficult for students to attain.

From an instructional point of view, "emphasis" may mean that more time in class is devoted to addressing this outcome and assuring that it is attained. With respect to assessment, "emphasis" may mean assigning more weight to test items related to a particular outcome or including a larger number of items for the outcome. It may also mean establishing a more stringent standard of acceptable performance for tests of these outcomes. For example, students who do poorly on test items dealing with dates and places of battles in a war may be allowed to advance to a new unit, whereas students who do poorly on items dealing with the political and social consequences of a war may be required to do remedial work before beginning a new unit.

Using Tests to Adapt Expected Outcomes to the Learning Backgrounds and Abilities of Students

What and how well students are expected to learn depends in large measure on the knowledge, skills, and aptitudes they bring to a learning situation. Students with limited backgrounds or abilities obviously need to begin work on more basic material than well informed or talented students. Expected learning outcomes appropriate for some classes or groups of students within a class may be inappropriate for others.

Tests often are helpful in determining the appropriateness of an intended outcome for a particular class or group of students. In most individualized instructional programs, for example, "locator" tests are provided to establish which objectives a student beginning the program should work toward. The wider the range of student backgrounds and ability levels in a class, the more varied will be the objectives that learners pursue.

Most teachers have neither the time nor the resources, however, to "customize" objectives for each and every student. Furthermore, in many classes, such as accelerated, advanced placement, or remedial classes, differences among students are not great. In these classes, a pretest or comparable assessment procedures may indicate the general appropriateness of a set of expected outcomes. When differences among learners are great, pretests, locator tests, or other diagnostic devices may suggest the need to adapt expectations for a particular group or groups of students.

Preparing Outcome Statements for Use by Students

When students know what and how well they are expected to learn, they are in a stronger position to manage and monitor their own learning progress than when learning expectations are vague or unstated. Teachers may inform students verbally about learning expectations. However, it may also be helpful to provide students with a written list of the learning outcomes expected from a unit or a set of lessons. This list obviously should be written in language appropriate for students. Moreover, the list should be reasonably brief, since students are likely to feel bogged down with lists that exceed a single page or so. An example of a list of outcome statements prepared for use by students is presented below.

A List of Outcomes Expected From a Unit on the Nature of Scientific Thinking

Prepared for Use by Students

	Degree of EMPHASIS we will place on this in class and on tests		
What you should KNOW:	**High**	**Moderate**	**Low**
1. the difference between questions that can be studied through scientific means and those that can't		X	
2. the difference between a scientific and nonscientific approach to studying questions		X	
3. the way in which a scientist's frame of reference influences his or her choice of questions to examine		X	
4. the meaning of terms like "hypothesis," "law," "theory" and "model"			X
5. the role that experiments play in scientific research			X
What you should be able to DO:			
1. design simple experiments to find out which of several factors, e.g., light, moisture, or temperature, influence the behavior of a given animal or plant the most	X		
2. decide whether a particular hypothesis is or is not supported by a given theory	X		
3. decide whether the results of a particular study support or put into question a given hypothesis	X		

RESOURCES

Chapter 6

Baker, E.L. and Popham, W.J. *Expanding dimensions of instructional objectives.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Gronlund, N.E. *Stating behavioral objectives for classroom instruction.* New York: Macmillan, 1970.

Hambleton, R.K. Advances in criterion-referenced testing technology. In C. Reynolds and T. Gutkin (Eds.), *Handbook of school psychology.* New York: Wiley, 1982.

Hannah, L. S., & Michaelis, J. N. *A comprehensive framework for instructional objectives.* Reading, Mass.: Addison Wesley, 1977.

Mager, R.F. *Preparing instructional objectives.* Palo Alto, Ca.: Fearon Publishers, 1962.

Popham, W.J. Domain specification strategies. In R.A. Berk (Ed.). *Criterion-referenced measurement: The state of the art.* Baltimore: The Johns Hopkins University Press, 1980.

SBB6

CHAPTER 7

MATCHING TEACHING TO EXPECTED LEARNING OUTCOMES

Introduction

In Chapter 3, which dealt with the context of instruction, it was noted that a number of models of effective teaching have been developed that provide frameworks for organizing and managing lessons. These models do not, however, indicate how instructional techniques may be varied to accommodate different kinds of intended learning outcomes.

The purpose of this chapter is to provide a basis for matching instruction to particular types of expected outcomes. Guidelines are presented for promoting students' learning of facts, concepts, principles, and procedures at various levels of attainment. The chapter rests on the assumption that students gain knowledge of content and demonstrate that they can **remember** it before they are taught to **use** content or expected to **refine** it. It is suggested that teaching strategies appropriate for achieving outcomes at the use-level build upon strategies appropriate for achieving outcomes at the remember-level. Likewise, strategies for promoting content refinement build upon strategies for promoting content use.

It is further assumed that one of the most important ways in which instructional practices vary by type of outcome expected is in the clarification and structuring of content to be taught, e.g., grouping facts into meaningful units, or identifying the attributes of a concept. The teaching practices described in the chapter thus include planning activities as well as modes of interacting with students.

Teaching Facts

Three instructional practices that promote students' learning of factual material are described in this section. Since facts in themselves cannot be used, these practices focus only on acquiring and remembering facts.

1. Organize facts in relation to larger units of knowledge

Isolated facts often are meaningless and difficult to remember. The individual fact, "George Washington was the first President of the United States," for example, is meaningful only in the context of other facts about the formation of the American Republic. To be fully understood and remembered, facts should be organized by topic, theme, or comparable category. For example, facts about the effects of pesticides might be grouped under the categories of "effects on agricultural production," "effects on the natural environment," and "effects on human health."

2. Indicate to students at the beginning of a lesson or assignment what categories of facts they are expected to remember

Students attend more effectively to a presentation when they know what kind of information they are to focus on. Here is an example of a statement designed to structure students' attention: "When viewing this film on the political aftermath of World War I, pay particular attention to the reasons offered for Congressional opposition to the League of Nations. You'll be expected to identify three of these reasons in discussion following the film."

3. Consider the use of special memory aids

Memory aids, i.e., "mnemonic devices," might be used to help students remember factual material. Memory aids include songs, rhymes, diagrams, and images that students associate with the information to be recalled. The use of the term SOH CAH TOA, for example, to stand for Sine A = opposite/hypotenuse; Cosine A = adjacent/hypotenuse; and Tangent A = opposite/adjacent, has proven helpful to many students of mathematics. Another well-known mnemonic device is "Roy G. Biv," which is used to represent the order of colors in a spectrum or rainbow: Red, orange, yellow, green, blue, indigo, and violet. Encouraging students to invent their own mnemonic devices may also enhance students' ability to remember what has been studied.

Teaching Concepts

The first six practices described in this section aim to foster students' knowledge of concepts and their ability to remember them. These six practices form the foundation for several additional practices that are described, which aim to facilitate concept use and concept refinement.

To Promote the Acquisition of Concepts at the Remember Level

1. Check whether students have attained prerequisite concepts

Concepts often are defined in terms of other concepts. For example, in chemistry a "precipitate" is defined as an insoluble substance which forms from a solution. If students do not know the concepts "insoluble," "substance," and "solution," they would not be prepared to learn the concept of precipitate. As another example, the concept of "anomie" in sociology is defined as a state in which social norms are weak or lacking. Understanding the concept of "norm" and the functions that norms serve is necessary to understand anomie. Identifying and assessing prerequisite concepts is essential in preparing to teach a new concept.

2. Identify each of the defining attributes of the concept

The defining attributes of a concept are those properties that give it a distinctive identity. For example, Klausmeier (1975, p. 286) lists the following defining attributes of the concept "equilateral triangle":

- plane
- simple
- closed figure
- three
- equal length

An equilateral triangle lacking any one of these attributes would not be an equilateral triangle.

A good definition of a concept includes all of its defining attributes. A definition is overly general if any of its defining attributes are missing, e.g., defining a bureaucracy as a "large organization," without including the attributes of specialized functions and formalized regulations and procedures.

3. Identify significant irrelevant attributes of the concept

An irrelevant attribute of a concept is a quality that is associated with the concept, but does not define it. For example, "size," and "orientation on the page" are irrelevant attributes of the concept equilateral triangle. As another example, irrelevant attributes of the concept "noun" include "number of syllables in the word," "initial letters of the word" and "manner in which the word is written" (Klausmeier, 1975).

A concept can vary in its irrelevant attributes without losing its identity. An automobile is an automobile, for example, whether it's a Toyota, Ford, or Chevrolet. The make of a car is an irrelevant attribute when classifying a vehicle as a car or as something else.

4. Construct or select examples and nonexamples of the concept

Examples used to teach a concept should highlight its defining attributes and clarify its irrelevant attributes. For instance, when teaching the concept of equilateral triangle, a good way to show the meaning of the defining attribute, "closed figure," would be to contrast an example of a closed figure with a nonexample of the attribute, i.e., an open figure. Similarly, to convey the meaning of the attribute, "three-sided figure," one might contrast a three-sided figure with a four or five-sided one. Examples that vary according to each of the irrelevant attributes identified also should be selected. For instance, equilateral triangles of different size and with differing orientations on a page should be created to teach the concept of equilateral triangle.

Time and resources are not always available to build examples and nonexamples for each irrelevant and defining attribute. However, some examples and nonexamples that are quite similar to each other, and some that are quite different from each other, should be constructed.

5. Present the attributes of the concept or guide students to discover them

The most direct way to teach a concept is to present its defining and irrelevant attributes followed by examples and nonexamples illustrating these attributes. There is ample research indicating that this direct approach, often referred to as an "expository approach," is a highly efficient teaching strategy.

By contrast, a discovery approach to teaching concepts may also be used. In a discovery approach, the teacher presents examples and nonexamples of the concept, but does not state the attributes of the concept. Instead, through careful study of the examples and nonexamples, guided by teacher questions and probes, students are expected to discern what the examples have in common and how they differ from the nonexample. In this approach, students derive the defining and irrelevant attributes. For example, to teach the concept of democracy through discovery, students might be asked to read descriptions of dictatorships, monarchies, and oligarchies, on the one hand, and descriptions of various forms of democratic government on the other hand. They then would be asked to determine why the sets of examples were grouped as they were, i.e., what characteristics were evident in each description of a democracy and were not evident in the descriptions of non-democratic governments. Students' statements regarding the basis of the classifications might then be compared to the definition of democracy stated in a text or one previously created by the teacher. Similarities and differences between the student-constructed definition and expert constructed ones might be discussed and a synthesis of the two created.

Proponents of the discovery method maintain that learning to acquire concepts through this method lays a stronger foundation for students' subsequent use of concepts than does the expository method. It may be, however, that the relative effectiveness of the discovery or expository method depends on the background and characteristics of the students with whom a teacher is working. For example, a direct method may be more effective than a discovery method among students with low ability or limited preparation in a learning area.

6. Check that students can restate or paraphrase the defining attributes of the concept and a few examples and nonexamples of the concept

While remembering definitions and examples usually is not the end-point of concept learning, it generally facilitates concept use. A student who cannot identify the attributes of "drosophilia," for example, probably would have difficulty distinguishing between drosophilia and closely related organisms.

Knowing the precise definition of a concept also is important when the concept is a basis for learning other concepts and principles. For example, a student might learn the sociological concepts of "role," "norm," and "function" as a basis for understanding the larger concept of "institution." If students have a flawed understanding of a concept, their understanding of related concepts also will be flawed.

To Promote the Acquisition of Concepts at the Use-Level

7. Provide practice in using the concept

To use a concept is to recognize newly encountered objects, events, or symbols as examples or nonexamples of the concept. For example, to use the concept of "chemical change" students would recognize instances of chemical change and distinguish them from instances of "physical change."

Practice in concept use involves generalizing to new examples and nonexamples of the concept. Initially, examples and nonexamples furnished for practice should be similar to those furnished through instructional presentations.

As students demonstrate skill in classifying examples that are similar to those previously encountered, the complexity of the context in which examples are presented may be increased. For instance, the concept of discrimination may have been illustrated initially through several brief case studies involving legal discrimination in housing, in public facilities, and in educational institutions. Practice in applying the concept, however, might involve the students in reading relatively long sections of autobiographies of well known minority-group members and identifying the various covert, as well as overt forms of discrimination that the authors encountered. In a lesson on the concept of irony, students might initially be presented with statements that obviously express irony and those that just as obviously do not. The teacher might then ask students to identify ironic elements that are subtly expressed in a short story. This kind of practice extends the meaning of a concept as well as stretches the limits of students' classification skill.

8. Ask students to generate examples of the concept

Students strengthen their skill in using a concept by finding examples of it from their own lives, or from other sources not explicitly provided in class. For instance, students might be asked to identify examples of discrimination in their own social groups or activities.

To Promote the Acquisition of Concepts at the Refine-Level

9. Provide students with examples, or ask them to generate examples, that challenge established definitions

Probably the most common circumstance that gives rise to conceptual refinement is encountering an example that neither clearly belongs to the class of objects, events, symbols, etc., named by the concept, nor clearly does not belong. These borderline examples represent a puzzle. To resolve the puzzle, a new definition of a concept may need to be fashioned.

To cite a simple example, a student may define a "dog" as a "4-legged animal with hair." When other 4-legged animals with hair are encountered, the student is forced to revise the definition, or face a contradiction. This process of revising definitions in view of new evidence is the kind of process stimulated by discovery teaching.

Teaching Principles

The first three guidelines presented in this section are useful for developing students' knowledge of principles. The subsequent guidelines focus on promoting skill in using and refining principles.

To Promote the Acquisition of Principles at the Remember-Level

1. Check whether students have attained prerequisite concepts

Principles specify a relationship between concepts. The most common types of relationships expressed through principles deal with causes and effects (A causes B); correlations (A is associated with B); necessary conditions (A enables B - e.g., "Individuals must care about themselves before they can care about others"); and ethical obligations, (e.g., "Do unto others as you would have them do unto you.")

In order to learn a principle, students need to have acquired the individual concepts that make it up. For example, a student who does not understand the concepts of "like charges" and "unlike charges" would not be able to understand a basic law of electrostatics that like charges repel each other and unlike charges attract each other.

2. Make clear the nature of the relationship specified in the principle

This can be done through either an expository approach or a discovery approach. In the expository approach the teacher provides a complete explanation of the relationship contained in the principle. This frequently involves the use of models, diagrams, analogies, e.g., using tinker toys to show principles of chemical bonding.

In the discovery approach the teacher presents, enacts, or involves students in situations that illustrate the principle. Students then *derive* the principle from their observation or experience. For example, a teacher might ask students to view several video tapes of young children interacting with each other. Some of the tapes might show older children repeatedly taking toys and candy from younger playmates, who in turn engage in various aggressive acts against objects at their disposal, e.g., banging dolls against the wall, tearing pages from books, etc. The other tapes show older and younger children playing cooperatively with no evidence of aggressive behavior on the younger children's part. After viewing the tapes, students might be asked to venture an explanation as to why the younger children behaved differently in different tapes. If students responded that the older children treated the younger ones better in some situations than in others, the teacher might then ask *why* that would lead to the behaviors observed. Through this kind of questioning, and through observation of similar patterns of interaction in different settings, students may arrive at the principle that "repeated frustration leads to aggression," or some comparable general relationship. In an expository approach, concrete illustrations come after the teacher's explanation of the principle. In discovery, concrete illustrations serve as a stimulus for deriving the principle.

3. Check to see that students can restate or paraphrase the principle

Knowing the correct formulation of a principle generally facilitates application of the principle. If students remember only part of the principle, or if they substitute or add extraneous terms to the formulation, they are not likely to use the principle as effectively as when they know it precisely.

To Promote the Acquisition of Principles at the Use-Level

4. Provide an appropriate range of practice situations

Students need practice and feedback in using principles across an appropriate range of situations. The appropriate range of practice situations depends in large part on the generality/specificity of the principles to be mastered.

The more general a principle the wider is the range of situations to which it relates and the greater is the degree of interpretation needed to apply it. The broad ethical principle, "Always treat people as ends, never as means," for example, has relevance in countless social situations. But considerable judgment is needed to assess the implications of the principle for a particular moral question or decision. In contrast, the rule "Proper nouns are capitalized" relates to a much narrower set of circumstances, and can be applied with virtually no deliberation or interpretation.

A larger and more diverse number of practice situations needs to be provided when teaching broad principles as compared to narrow ones. In addition, correction procedures for teaching broad principles are more likely to focus on the reasons underlying student errors or inadequacies than correction procedures for teaching narrow principles. Students who fail to apply the principle that all proper nouns are capitalized, for example, probably either have forgotten the principle, or don't know what a proper noun is. The causes of poor performance in applying this kind of principle would be simple to address. However, students who experience difficulty in applying Einstein's theory of relativity in explanations of various physical conditions and changes may harbor any number of misconceptions. Teachers would need to probe to uncover these misconceptions and to correct them.

5. Encourage students to find new situations in which the principle can be used

New situations in which to apply a principle may be encountered on television, in newspapers, in other classes, in day-to-day life, or in students' imaginations. Searching for new opportunities to apply a principle will enhance students' ability to use a principle independently.

To Promote the Acquisition of Principles at the Refine-Level

In mathematics, science, and technological fields, the vast majority of principles that students learn are well-established and not likely to require refinement. Principles of psychology, sociology, ethics, and those in other social science and humanities areas may be more amenable to refinement. For example, students might be given the opportunity to refine the psychological principle, "A person is more likely to act on a motive if he clearly understands it" (based on McClelland, 1965). Students may wish to qualify this statement by identifying conditions under which the principle does not hold, e.g., people may be less likely to act on violent or aggressive motives if they understood them.

Principles thus may be modified and improved in view of evidence that they are overstated, oversimplified, unclear, or in some other way deficient. One way to stimulate students to refine a principle is to present evidence of its limitations, e.g., evidence that a nation's rate of inflation is escalating much faster than an economic principle or set of principles would predict.

Teaching Procedures

The following guidelines apply to the teaching of procedures in general, whether they be procedures to guide thinking about problems and issues or procedures to guide performance on athletic, artistic, or manual tasks. The first four guidelines are used to teach students about procedures. These guidelines lay the basis for subsequent guidelines on promoting the use and refinement of procedures.

To Promote the Acquisition of Procedures at the Remember-Level

1. Check whether students have attained prerequisite concepts and principles

Knowledge of procedures depends on knowledge of concepts, and may also depend on knowledge of principles. As a simple illustration, a student would not be able to acquire procedures for driving a car without knowledge of the concepts of "road," "steering wheel," "dash," "brake," etc. As another example, procedures for repairing a complex machine generally depend on the user's knowledge of principles underlying the machine's operation. As a departure point for teaching a procedure, it generally is a good idea to assess students' understanding of the concepts and principles that are embedded in it.

2. Indicate the context in which the procedure is used

Students sometimes learn to use a procedure without knowing when and why it is to be used. For example, students may learn how to diagram the grammatical structure of sentences or the conceptual relations contained in reading passages. However, they may not know the purposes to be served by these procedures or the conditions under which they are to be used. A description of the context, or varying contexts, in which the procedure can be used facilitates meaningful, as opposed to rote, learning of principles.

3. Demonstrate how the procedure is used and explain each step along the way

Students learn a good deal from observing a model of effective performance. Modeling commonly is done when teaching physical skills, like shooting a layup in basketball, using a particular dismount in gymnastics, and using woodworking tools in industrial arts. Modeling is less commonly used when teaching procedures that guide and enhance thinking, like problem solving and inquiry procedures. However, these procedures also can be modeled by the teacher, by competent students, or by other skilled persons. For example, a teacher can show the steps involved in summarizing ideas presented in a reading passage (Brown et al., 1983), in studying scientific text (Larkin & Reif, 1976), or in analyzing controversial social issues (Oliver & Shaver, 1966).

To Promote the Acquisition of Procedures at the Use-Level

4. Provide an appropriate range of practice situations

Students need to receive practice and feedback in using procedures in an appropriate range of situations. The appropriate range of practice situations varies according to the breadth of the procedures to be mastered. The more general the procedure, the wider is the range of situations in which it applies, and the greater is the level of interpretation and judgment required to use it. For example, procedures for typing a memorandum are narrow in scope and can be followed without extensive reflection. Procedures for solving story problems in mathematics usually are abstract, cover a variety of classes of story problems, and call for extensive independent analysis on the students' part. When broad procedures are taught, students need practice in adapting procedures to specific contexts. Rules for varying procedures by context may need to be created to assist students in transferring their procedural knowledge to unfamiliar situations.

To Promote the Acquisition of Procedures at the Refine-Level

5. Encourage students to modify and extend procedures to enhance their utility and effectiveness

Students may personalize a procedure by adapting it to their own styles and needs, e.g., using a two-handed backhand stroke in tennis after finding limitations in the one-handed procedure that was taught. Students may also propose that a procedure be refined generally, e.g., they may suggest a new way to monitor the extent and quality of participation in class discussions in view of their experience with the procedure.

Though there doesn't appear to be any direct way of teaching students how to refine procedures, other than through practice with feedback as to quality of performance, teachers may generally encourage students to improve established practices and may provide extensive opportunities for them to do so. They may point out specific aspects of a procedure that students may wish to modify, e.g., the sequence through which a particular set of steps are to be taken, or the time to be allotted for a specific operation. When a student is successful in improving a procedure, the accomplishment might be shared with fellow students, and the process underlying the improvement might be explored.

REFERENCES

Chapter 7

Brown, A. L., Day, J. D., & Jones, R. S. The development of plans for summarizing texts. Technical Report No. 268. Champaign, Ill.: Center for the study of Reading, University of Illinois, 1983.

Klausmeier, H. J. Learning and human abilities. New York: Harper and Row, 1975.

Larkin, J. A., & Reif, F. Analysis and teaching of a general skill for studying scientific text. Journal of Educational Psychology, 1976, 68, 431-440.

McClelland, D.C.. Toward a theory of motive acquisition. American Psychologist, 1965, 20, 321-333.

Oliver, D. W., & Shaver, J. P. Teaching public issues in the high school. Boston: Houghton Mifflin, 1966.

RELATED RESOURCES

Cooper, J.M. & Associates. Classroom teaching skills: A handbook. Lexington, Mass.: D.C. Heath, 1977.

Gage, N.L. (Ed.) The psychology of teaching methods. (Seventh-fifth yearbook of the National Society for the Study of Education, Part I). Chicago: University of Chicago Press, 1976.

Gagne, R.M. & Briggs, L.J. Principles of instructional design. New York: Holt, Rinehart, and Winston, 1974.

Green, T.F. The activities of teaching. New York: McGraw Hill, 1971.

Joyce, B., & Weil, M. Models of teaching. Englewood Cliffs, N.J.: Prentice-Hall, 1982.

Klausmeier, H.J., Lipham, J. & Daresh, J.C. The renewal and improvement of secondary education: Concepts and practices. Lanham, Md.: University Press of America, 1983.

Merrill, M.D. & Tennyson, R.D. Teaching concepts: An instructional design guide. Englewood Cliffs, NJ.: Educational Technology Publications, 1977.

Reigeluth, C.M. (Ed.) Instructional design theory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

CHAPTER 8

MATCHING TEST ITEMS AND ASSESSMENT PROCEDURES TO EXPECTED LEARNING OUTCOMES

Introduction

Just as approaches to instruction need to be matched to the learning outcomes desired from instruction, tests also need to be matched to the kind of learning outcomes to be assessed. Measurement specialists often refer to the match between expected outcomes and tests as "content validity." The purpose of this chapter is to provide a basis for selecting the approach to testing that is most appropriate for assessing particular types of learning outcomes. Although most test formats can be used to assess a variety of learning outcomes, particular formats are better for assessing some types of outcomes than others.

In addition to assuring that one's approach to assessment is appropriate to the outcomes to be assessed, the time required for assessment (on the part of both students and teachers), and the utility of information obtained, also need to be considered. The test format selected for a particular assessment task should be the most efficient to administer and score given the particular kind of learning outcome to be assessed, the particular kind or level of information desired and the uses to be made of this information.

Selecting a test format that takes all of these matters into account, and provides for the best match possible, requires considerable knowledge as well as careful thought. This chapter is designed to help teachers in this complex level of decision making.

The chapter has two major sections. The first describes the use of objective and interpretive tests in assessing **facts, concepts**, and **principles**. The second describes a variety of test formats that can be used in assessing students' knowledge of **procedures** and their skill in applying and refining them. Procedures are defined broadly in the handbook to include processes, strategies, and techniques. To accommodate the wide range of procedures that can be assessed, e.g., problem solving procedures, artistic techniques, and athletic routines, a separate section has been devoted to assessing procedures.

Assessing Student Mastery of Facts, Concepts, and Principles

The types of formats typically used to assess the acquisition of facts, concepts and principles, at various levels of mastery, are identified in this section. The relationships between learning outcomes and assessment procedures to be dealt with in this section of the chapter can be indicated as follows:

	Remember-Level	**Use-Level**	**Refine-Level**
Facts	Objective tests	NA	NA
Concepts	Objective tests	Objective tests or Essays & Product-related tests	**Commonly used:** Essays & Product-related tests: **Sometimes used:** Objective tests
Principles	Objective tests	Objective tests or Essays & Product-related tests	**Commonly used:** Essays & Product-related tests **Sometimes used:** Objective tests

Assessing Facts, Concepts and Principles at the Remember-Level

Objective tests provide the most direct and efficient means of assessing students' ability to remember facts, concepts, and principles. This is not to deny that essays and other student products or direct observations can provide evidence on what students remember, but they tend to be inefficient ways of doing so. For example, a teacher who observes a student running with a football in the wrong direction on the field may infer that the student can't remember, or never learned to begin with, some of the basic facts of the game. When the primary focus of assessment is on students' performance at the remember-level, however, objective tests that focus squarely on knowledge possessed or retained after a learning experience are probably the best and most efficient mode of assessment.

Both forms of objective tests, response-selection and response-completion, are useful in assessing students' performance at the remember-level. Response-selection items generally are less demanding as a measure of knowledge than response-completion items, since they require students only to recognize something, and response-completion items require students to state something ("name," "list," "supply," etc.), or paraphrase it. Both, however, represent acceptable approaches to measurement at this level of mastery. Examples of response selection and response completion items that assess facts, concepts, and principles at the remember-level are presented below.

	Response Selection Items	**Response Completion Items**
Facts	Coal is a fossil fuel. *True False	Name a fossil fuel.
Concepts	According to your textbook, which of the following is the correct definition of the concept "social norm"? A. A law that most people obey *B. A shared expectation for behavior C. A fundamental belief D. A pattern of thought or action	Define the concept "social norm." Use your own words.
Principles	The law of conservation of energy states that the total energy of the universe remains constant. *True False	What does the law of conservation of energy state?

Assessing Concepts and Principles at the Use-Level

Both objective and essay tests can be used to assess students' ability to use concepts and principles. When appropriate applications of a concept or principle are to be offered by students, or when the applications to be suggested cannot be anticipated, an essay or product-related test will usually be preferred.

Regardless of the approach selected, **it is assumed that the examples or problems presented to students are different from the ones used for instructional purposes.** Otherwise, the items would merely be assessing performance at the remember-level. Test items illustrating the use of both objective and non-objective approaches to assessing students' ability to apply (use) knowledge are presented on the following page:

TABLE 8-1

Illustrative Items for Assessing Concepts and Principles at the Use Level

	Objective Test Items	Essay Questions
Concepts	(to assess students' ability to apply the concepts of "fact" and "opinion") Each of the 5 statements below is either a fact or an opinion. Read each statement. If the statement is a fact, place an "F" in the space beside it. If the statement is an opinion, place an "O" in the space beside it. ____ 1. It is too cold and rainy to play football. ____ 2. German shepherds make the best family pets. ____ 3. Los Angeles is the largest city in California. ____ 4. Despite Diamond Jim Brady's popularity, his life didn't amount to much. ____ 5. About eight percent of the U.S. work force in 1960 was employed in agriculture, as compared to 33 percent in 1910.	Do you think the school store is an example of the concept "monopoly"? Give reasons to support your position.
Principles	In the space beside each statement of fact listed below, write the letter of the accompanying principle which is most useful in explaining it. **Statements of fact:** __ 1) Shears used to cut sheet metal have long handles. __ 2) The force exerted on a brake by the driver's foot is much less than exerted on the brake drums. __ 3) A rocket can propel itself in a vacuum. __ 4) If a rapidly rotating grindstone bursts, the fragments fly outward in straight lines. **Explanatory principles:** A. Force is equal to mass times acceleration. B. The momentum of a body tends to remain constant. C. The moment or turning effect of a force is proportional to its distance from the axis of rotation. D. Friction exists between bodies in contact and moving with respect to one another. E. The sum of kinetic and potential energies in an isolated system is a constant.	We have studied in some depth the economic theories of Milton Friedman and John Kenneth Galbraith. Based on your understanding of these theories, how do you think Friedman and Galbraith would respond to an increase tariff on Japanese goods imported to this county?

(adapted from Bloom et al., 1981, p. 240)

BEST COPY AVAILABLE

Assessing Concepts and Principles at the Refine-Level

It is difficult to use objective tests to assess students' ability to refine concepts or principles. Essay tests commonly are used to assess students' learning at this level of mastery. The items appearing below illustrate the form that essay questions of this kind take.

Concepts

We have seen in this unit that Freud developed the concept of "id" to represent what he viewed as the primitive sexual and aggressive instincts in the human personality. How would you extend or revise Freud's concept of "id" to accommodate recent claims that all of us are influenced to one degree or another by our "reptilean brain," which sends us highly non-rational messages.

Principles

During the first quarter of the course we developed a model of political revolution based on an analysis of the American, French, and Russian Revolutions. The model specified a) the conditions that need to be present for a revolution to succeed; b) the stages through which revolutions progress; and c) the structures that need to be established to sustain the changes created by the revolution.

Now that we have studied the Chinese, Cuban, and Iranian Revolutions, in what ways, if any, would you revise or extend our model of revolution? Provide a rationale for each proposed change.

Students' skill in refining concepts or principles can be assessed objectively when a single correct or best refinement can be identified. For example, the item above about the model of political revolution could be converted into an objective item if a teacher were certain that the model needed to be revised in a particular way to square with evidence the class had studied. He or she might present this revision along with three or four distractors in a multiple choice item. However, it is anticipated that in most cases the intent of items measuring performance at the refine level will be to see what creative adaptations or extensions of content students can produce, as distinct from merely selecting.

Assessing Student Mastery of Procedures

A separate section has been reserved for a discussion of approaches to assessing student mastery of procedures in view of the wide variety of procedures that teachers and students typically deal with. For ease of analysis, the various types of procedures commonly encountered in classroom situations are grouped into two broad categories: "Reasoning processes and problem solving procedures" and "Procedures that guide performance on artistic, athletic, interpersonal and manual tasks." This division has been made because testing practices appropriate for assessing these two types of procedures at the use-level and refine-level are somewhat different.

The chapter focuses on the relationships between the learning outcomes and assessment procedures indicated below.

	Remember-Level	**Use-Level**	**Refine-Level**
Reasoning Processes and Problem Solving Procedures	Objective tests	**Commonly used:** Objective tests or Essays & Product-related tests **Sometimes used:** Observation	**Commonly used:** Objective tests or Essays & Product-related tests **Sometimes used:** Observation
Procedures that Guide Performance on Artistic, Athletic, Interpersonal and Manual Tasks	Objective tests	**Commonly used:** Observation or Essays & Product-related tests **Sometimes used:** Objective tests	**Commonly used:** Observation or Essays & Product-related tests **Sometimes used:** Objective tests

Assessing Procedures at the Remember-Level

As in the case of facts, concepts, and principals, objective tests also are widely used to assess students' ability to recognize, restate, or describe procedures. Here are some examples of how objective test items can be used to assess student knowledge of procedure.

Example 1: Procedures to Guide Problem Solving

> According to Polya, there are four main steps in solving a mathematical problem. List these.

Example 2: Procedures to Guide Athletic Performance

In volleyball, a new person assumes the position of "server" every time a point is made.

True *False

Example 3: Procedures to Guide Manual Performance

Describe the steps that should be followed in changing a flat tire on a car.

Assessing Procedures at the Use-Level

In this handbook it is assumed that there is a close relationship between procedures and what commonly are called "skills." A procedure is an established way of doing something. When students learn to use a procedure successfully, it may be said that they have learned a skill. Assessing procedures at the use level is comparable to assessing skills. Approaches to assessing students' skill in using two broad classes of procedures are described in the paragraphs that follow.

1. Procedures Guiding Reasoning and Problem Solving

Both objective and essay or product-related tests can be used to assess students' ability to reason and use problem solving procedures. Teachers commonly infer from results on these tests that students have or have not used particular procedures, or have used them efficiently or inefficiently. For example, a teacher might conclude that a student has developed sophisticated strategies for literary analysis if he or she consistently demonstrates on essay tests an ability to identify the underlying themes of literary works. The teacher might not observe the application of these strategies directly, but the quality of the student's insights would suggest that such strategies are being used.

Objective and non-objective tests can be designed, however, to provide more direct information on the procedures students use to reach solutions or produce products.

An example of an objective test item and an essay-type question that provide information on the thinking processes students use to arrive at solutions to problems are presented in Table 8-2. Note that the items do not merely ask students to produce solutions. Rather, they ask them to choose or describe appropriate problem solving procedures.

When time and resources permit, a teacher may also make formal observations of students' reasoning processes. Of course, reasoning processes cannot be seen directly. But, it is possible to watch individual students very closely as they engage in a complex

TABLE 8-7

Illustrative Items for Assessing Skill in Selecting Appropriate Problem Solving Procedures

Objective Item

How would you decide which of the following farmers got the best yield of oats from his land?

Farmer	(1) Land Planted In Oats (acres)	(2) Total Oats Harvested (tons)
Mr. Davis	65	38
Mr. Jackson	75	75
Mr. Brown	38	53
Mr. Smith	50	40
Mr. North	25	30

Circle the letter of the correct answer.

A. Look at each number in the "land planted in oats" column and pick the farmer with the most land planted in oats.

B. Multiply each number in the first column by its number in the second column and pick the farmer with the largest answer.

C. Look at each number in the "total oats harvested" column and pick the farm with the most total oats harvested.

D. Divide each number in the second column by its number in the first column and pick the farmer with the largest answer.

(adapted from Ross & Maynes, 1981, p. 124)

Essay Question

Your school has just orbited its first living-lab station in space. You are on board as an environmental systems analyst. One day maintenance personnel report to you that the simple device you've constructed to measure atmospheric pressure indicates that, with respect to standard earth pressure of 101 kPa, the system atmosphere pressure is low and dropping.

List the steps you would take, and the order in which you would take them, to determine the cause of the problem. Indicate the reasons why you would take each of these steps in the particular sequence you propose.

(adapted from the Ontario Ministry of Education, 1981, p. 101) intellectual task and ask them to explain why they are approaching the task as they are. This is precisely the testing approach that school psychologists use when administering portions of an individual intelligence test. The psychologist not only notes how well a student completes a task, e.g., assembles blocks in a particular pattern, but how the child performs it, e.g., does he or she use a random, "trial and error" strategy, or a systematic one? Classroom conditions do not permit the kind of in-depth observations of student performance that are associated with individually administered psychological tests. However, teachers may use simplified observational procedures for assessing reasoning/problem solving processes when working with small groups of students, or when providing "extra help" to a student after school. This is illustrated by the checklist presented in Table 8-3 that a teacher could use to focus his or her observations of a student's planning strategies in solving mathematical problems.

TABLE 8-3

A Checklist For Observing Students' Use Of Planning Strategies In Math Problem Solving

The student clarifies a problem before working toward and implementing a solution. He or she:

	Yes	No
1. reads the problem several times	____	____
2. translates the problem into visual form, e.g., a graph, table, diagram, flowchart, etc.	____	____
3. lists the knowns and unknowns in the problem	____	____
4. breaks the problem into manageable parts	____	____

2. Procedures That Can Be Used in Assessing Performance on Artistic, Athletic, Interpersonal, and Manual Tasks

Observations are frequently made to assess students' use of particular processes and procedures called for in artistic, athletic, interpersonal, or manual tasks. For example, a teacher might wish to assess a student's skill in making informal introductions. To do this, one student might be asked to play the part of a new member of the class or a new person in the neighborhood, and another student might be requested to introduce him or her to classmates, friends, parents. A checklist that might be used to rate qualities of a student's introduction is presented in Table 8-4.

TABLE 8-4

An Illustrative Checklist for Assessing a Student's Skill in Making Informal Introductions

The student	Yes	No
. said the name of the person(s) being introduced in a clear and distinct voice	____	____
. conveyed something significant about the person(s) being introduced	____	____
. indicated how he or she knew the person(s) being introduced	____	____
. used "ice breakers" or other appropriate expressions to stimulate conversation between the person being introduced and others in the group	____	____

Product-related tests also are employed frequently to assess artistic, athletic, interpersonal, and manual skills. For example, a scale that a teacher might use to assess the quality of a student's technical drawing is presented in Table 8-5.

TABLE 8-5

An Illustrative Rating Scale for Evaluating a Technical Drawing

1	2	3
Unacceptable	**Acceptable**	**Distinguished**
Drawing does not match specifications; it is inaccurate, illegible, messy, or incomplete	Drawing meets all specified requirements; it is legible and neat; it is accurate, complete, and provides all needed information	Superior drawing quality; all parts labeled clearly; matches all specifications; provides additional information and detail

(Adapted from Priestly, 1982, p. 141)

Objective test items conceivably also could be used to obtain information on artistic, interpersonal, and related skills. An item may present students with a complex interpersonal dilemma, for example, and ask students to select the most sensitive or ethical response to the dilemma. Objective test items, however, are not widely used to assess skills in areas like interpersonal communication, performing arts, or physical education.

Assessing Procedures at the Refine Level

Once students have learned to use a procedure, teachers may wish to assess students' ability to adapt or extend it. In many learning areas, however, particularly in the physical and natural sciences and in mathematics, the procedures students learn are well established and are not in need of revision or refinement. Only when there is a clear potential for students to improve upon a procedure that has been taught is it reasonable to assess procedures at the refine level. Two illustrative items that could be used to assess students' skill in refining procedures are shown in Table 8-6.

TABLE 8-6

Illustrative Items for Assessing Skills in Refining Procedures

Example 1: Problem Solving Procedures

We have spent two weeks studying and applying the STP (Situation-Target-Proposal) Model of Problem Analysis. However, a number of students have indicated that the model doesn't seem to work well in certain situations.

Your task is to find a way of making this problem solving model more generally effective. First indicate the limitations of the model as it is currently conceived. Then propose changes in the model and provide a rationale for each change proposed. Indicate also how you plan on testing the effects of your proposed changes.

Example 2: Gymnastic Skill

Create a routine on the parallel bars that is at a higher level of difficulty, i.e., demands more skill, than the routine you recently have mastered. The routine should reflect your conception of what is exciting and graceful about the parallel bars, and what you as a gymnast are capable of achieving. Avoid copying routines you have observed others performing.

Your routine will be evaluated in terms of its level of difficulty and its originality.

REFERENCES

Chapter 8

Bloom, B. S., Madaus, G. F., & Hastings, J. T. Evaluation to improve learning. New York: McGraw Hill, 1981.

Ontario Ministry of Education. Assessment Instrument Pool in Chemistry (I, Senior Division). Toronto, Ontario, 1981.

Ross, J.A. & Maynes, F. Geographic thinking skills. Toronto, Canada: Ontario Institute for Studies in Education, 1983.

RELATED RESOURCES

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Gronlund, N. E. Measurement and evaluation in teaching. (4th ed.). New York: Macmillan, 1981.

Mehrens, W. A. & Lehmann, I. J. Measurement and evaluation in education and psychology, 3rd ed. New York: Holt, Rinehart and Winston, 1984.

TenBrink, T. D. Evaluation: A Practical guide for teachers. New York: McGraw-Hill, 1974.

SBB8

CHAPTER 9

ASSURING QUALITY IN TEST ITEMS AND ASSESSMENT PROCEDURES

Introduction

Selecting a test format that is appropriate for a particular learning outcome is only one step in the process of test development. An item or assessment procedure may be constructed in an appropriate format, but nonetheless contain serious flaws. In this chapter, guidelines are provided for assuring quality in test items and assessment procedures. The chapter is divided into three parts. The first part deals with **objective tests**; the second deals with **essays and product-related tests**; and the third deals with **observation of student performance.**

Objective Tests

Procedures for writing various types of objective test items are described in this section of the chapter. Item types include multiple choice, alternative response, matching, fill-in-the-blank, and short answers or exercises.

Multiple Choice Items

Multiple choice items are highly versatile. They are used to assess a wide variety of learning outcomes, including those that call for the application or refinement of knowledge.

A multiple choice item consists of two parts: a **stem**, which defines the problem with which the item is concerned, and **alternatives**, the various responses to the stem from which the student must choose. One alternative is the correct or best response. The others are **distractors**, i.e., options that are plausible but incorrect or less adequate than the best response. Step-by-step guidelines for writing multiple choice items follow.*

1. **Be sure that the stem identifies clearly the problem the student is to consider.**

A stem should present a question or incomplete statement that can be responded to without seeing the alternatives.

Poor Stem	Improved Stem
The economy of Oregon....	On which of the following industries is the economy of Oregon most dependent?

2. **When the stem is in the form of an incomplete statement, the alternatives should finish the statement.**

When the alternatives finish an incomplete statement, the task of selecting an alternative can be carried out more efficiently, without having to read a stem over and over:

---

*The guidelines presented in this chapter are adapted from Bloom, Madaus and Hastings (1981); Brown (1981); Hambleton and Eignor (1978); Hopkins and Antes (1979); and TenBrink (1974).

Poor Item	Improved Item
The _____ is a device to measure specific gravity of liquids.	A device used to measure specific gravity of liquids is called a:
A. dosimeter	A. dosimeter
B. barometer	B. barometer
C. hydrometer	C. hydrometer
D. manometer	D. manometer
	(Hopkins and Antes, 1979, p. 112)

3. **Include words in the stem that otherwise would have to be repeated in each alternative.**

This makes the item less wordy and easier to respond to.

Poor Item	Improved Item
_____ Which of the following best summarizes Walt Whitman's "I Hear America Singing?"	_____ In Walt Whitman's "I Hear America Singing," the greatness of America is summarized by
A. The greatness of America is in its industrial strength, combined with an adequate labor force.	A. its industrial strength combined with an adequate labor force
B. The greatness of America is in the recognition of the individual's abilities.	B. the recognition of the individual's abilities.
C. The greatness of America is the individual's working independently and yet harmoniously.	C. individuals working independently and yet harmoniously.
D. The greatness of America is in each individual's appropriate work choice.	D. each individual's choice of work.
	(adapted from Hopkins & Antes, 1979, p. 113)

4. **Avoid using negative words in the stem. If a negative word must be used, underline or capitalize all letters of the word.**

Students usually find negatively worded stems confusing, since they are accustomed to choosing the correct, rather than the incorrect, answer.

Poor Item	Improved Item
Which of the following statements is not an example of irony?	All of the following statements are examples of irony EXCEPT

5. Be sure each distractor is plausible.

If a distractor is so obviously wrong that alert students would likely eliminate it even if they knew little about the problem defined by the item, the distractor should be replaced.

Poor Item	Improved Item
Between 1900 and 1910, most immigrants to the United States came from:	Between 1900 and 1910, most immigrants to the United States came from:
A. Germany	A Germany
B. Ireland	B. Ireland
C. Italy	C. Italy
D. Mozambique	D. Russia

6. Be sure that there is only one alternative that experts would consider correct.

This guideline needs little commentary. Care must be exercised that the answer designated as the correct response is in fact the correct or best option.

7. Be sure that the alternatives have the same grammatical structure, include similar terminology, and are about the same length.

Inconsistencies in grammar, terminology, and length can give students clues to the correct answer.

Inconsistencies in Grammar

Example: Frank Lloyd Wright was an

A. potter.
B. architect.
C. sculptor.
D. watercolorist.

(Bloom et al. 1981, p. 198)

Only option "B," the correct alternative, is properly modified by the "an" appearing in the item.

Inconsistencies in Terminology

Example: Which one of the following foods is the best source of quick energy?

1. Fat
2. Water
3. Glucose
4. Starch

(TenBrink, 1974, p. 372)

The correct answer, "Glucose," is the only alternative stated in technical language.

Inconsistencies in Length

Example: Which of the following statements is a fact?

A. The law is unfair.
B. The law would not have passed in 1982.
C. The law contains three new regulations governing non-profit organizations.
D. The law is long overdue.

The correct alternative, "C," stands out because of its length. Students may select longer options simply because they are longer. This often is an effective strategy, since there is a tendency for correct choices to be more qualified than incorrect choices and thus to be longer. Avoid the tendency to make the correct answer longer than the distractors.

8. When appropriate, arrange alternatives in a logical order.

Example: How many minutes is the length of an overtime period in college basketball?

A. 3
B. 4
* C. 5
D. 10

(Brown, 1981, p. 41)

9. Avoid positioning the correct alternative in a predictable order from item to item, e.g., the first or last alternative presented in each item.

A **checklist** for reviewing important aspects of multiple choice items is presented in Table 9-1.

TABLE 9-1

A Checklist for Preparing Multiple Choice Test Items

Stem

_____ (1) The stem clearly defines the problem with which the item is concerned.

_____ (2) If the stem is in the form of an incomplete statement, the alternatives finish the statement.

_____ (3) The stem includes words that otherwise would have to be repeated in each alternative.

_____ (4) The stem is free from negative words. If a negative word must be used, all letters of the word are underlined or capitalized.

Alternatives

_____ (5) Each distractor is plausible.

_____ (6) There is only one alternative that experts would consider correct.

_____ (7) The alternatives have the same grammatical structure, include similar terminology, and are about the same length.

_____ (8) Alternatives are arranged in a logical order, whenever appropriate.

_____ (9) The correct alternative is not positioned in a predictable order from item to item, e.g., the first or last alternative presented in each item.

Alternative Response Items

Alternative response items commonly are used to measure performance at the remember-level, though it also is possible to use them to assess students' ability to apply knowledge, e.g., they can be used to assess whether students can differentiate between a correct and an incorrect application of a particular problem solving procedure.

The main limitation of alternative response items is that they require the test maker to reduce response options to "black and white" categories, like "yes-no," "true-false," or "correct-incorrect." Many learning outcomes are not well assessed through this two-option response format. However, this item type, because it takes relatively little effort to write and answer, can provide considerable information to the teacher in a short amount of time. Guidelines for constructing alternative response items follow.

1. Make sure that the item can be answered unambiguously by one of the two response options.

Poor Items	Improved Items
T F Richard Nixon was one of the most corrupt presidents of the 20th century.	T F Richard Nixon was impeached by the Congress of the United States.
T F Being a high school teacher is highly stressful.	T F A higher percentage of high school teachers reported leaving the teaching profession because of job-related stress in the 1970's than in the 1960's.

2. Make sure the item contains a single important idea or piece of information.

If multiple ideas are included in an item, one idea could be true and another false, as in the following example furnished by Bloom and his associate (1981, p. 189):

Poor Item	Improved Item
Vitamins play a role in the regulation of metabolism, but do not provide energy. T F	Vitamins play a role in the regulation of of metabolism. T F

3. Avoid using words that give clues to the correct answer.

Words like "always," "never," "all," or "now" generally indicate an answer is false. Words like "usually," "by and large," or "often" generally indicate an answer is true.

Poor Item	Improved Item
All U.S. Senators serve for six years. T F	The normal term of office for U.S. Senators is six years. T F

4. Make sure the item is not so simple or obvious that it can be answered on the basis of common sense alone.

Poor Item	Improved Item
There is more than one way to express friendship. T F	The symbols used to express friendship in the Yurok tribe are more closely related to the friendship symbols of the Sioux than to the Apache. T F

5. Avoid negatively worded items.

Poor Item	Improved Item
"Loquacious" is not a synomyn for "garrulous." Yes No	"Loquacious" is a synonym for "garrulous." Yes No

6. Do not lift statements from a text.

Items that contain statements from a text students are using reinforce rote memorization.

7. Do not use trick items.

Poor Item	Improved Item
Abraham Lincoln took office as President of the United States in 1861.	Abraham Lincoln was elected President of the United States in 1860.

The item on the left above is true because while Lincoln was elected in 1860, he did not take office until 1981. The item appears to be designed to trap students, however, or to assess students' ability to detect traps. The item on the right assesses students' knowledge of more meaningful information.

8. Avoid the tendency to make true statements more qualified and thus longer than false statements.

As indicated earlier in the discussion of multiple choice items, there is a tendency for statements that are true to contain more qualifications than false statements, and therefore to be longer. The length of a statement, however, should not give a clue as to whether it is true or false.

9. Guard against having the correct answers to items fall into a predictable order, e.g., T.F.T.F.T.F.

A **checklist** for reviewing important aspects of alternative response items is presented in Table 9-2.

Matching Items

Matching items permit a large number of related facts, concepts, principles or procedures to be assessed in a small amount of space. Matching items generally are used to assess performance at the remember level. It is possible, however, to devise matching items to assess mastery at the use level, e.g., matching a set of scientific principles with a set of predictions which flow from them.

Matching items consist of two columns. Students are asked to match each item from one column with an item from the other column.

Example:

1. Select the term from the list at the right that is described by the phrase at the left. Each answer can be used only once.

(d) A body moving around the sun.
(a) A very large group of stars.
(c) A small body moving around a larger one.
(e) A shooting star.
(b) A small planet.

a. galaxy
b. asteroid
c. satellite
d. planet
e. meteor
f. rocket

(Brown, 1981, p. 54).

TABLE 9-2

A Checklist for Preparing Alternative Response Items

_____ (1) The item can be answered unambiguously by one of the two response options.

_____ (2) The item contains a single important idea or piece of information.

_____ (3) The item is free from words that give clues to the correct answer.

_____ (4) The item is not so simple or obvious that it can be answered on the basis of common sense alone.

_____ (5) The item consists of a positive statement.

_____ (6) The item has not been lifted from a text.

_____ (7) The item is free of trick phrases.

_____ (8) True statements are not more qualified and longer than false statements.

_____ (9) The correct answer does not fall into a routine pattern.

The entries in the left column generally are called premises, and those in the right column are called responses. Here is a set of guidelines for constructing matching items.

1. Be sure that each entry in a column refers to the same type of content.

When columns include different types of content, students may be able to determine appropriate responses based on a superficial understanding of the material. If a response column is to refer to "effects," for example, then all reponses should identify effects. If effects are mixed with other elements of content, some answers will clearly be incorrect. For instance, if students saw the following entry under a response column that listed effects, they would know it would be inappropriate: "A viola is a string instrument." Since this is not an effect, students could ignore the response.

Hopkins and Antes (1979, pp. 120-121) provide more subtle examples of a column with heterogeneous and homogeneous content:

Poor Item

(heterogeneous content)

Directions: For each name in column I, find the achievement in column II and place the letter on the line provided. Each achievement in column II is to be used one time or not at all.

	Column I	Column II
	Name	Achievement
___	1. Thomas Jefferson	A. Chief Justice
___	2. John Marshall	B. novels
___	3. Eugene O'Neill	C. cotton gin
___	4. Carl Sandburg	D. aviation
___	5. Eli Whitney	E. electric light bulb
___	6. Orville Wright	
		F. poems
		G. plays
		H. paintings

Improved Item

(homogeneous content)

Directions: For each inventor listed in column I, find his invention in column II and place the letter on the line provided. Each invention in column II is to be used one time or not at all.

	Column I	Column II
	Inventor	Invention
(D)	1. Thomas Edison	A. cotton gin
(E)	2. Robert Fulton	B. telegraph
(F)	3. Cyrus McCormick	C. air flight
(B)	4. Samuel Morse	D. electric light bulb
(A)	5. Eli Whitney	
(C)	6. Orville Wright	E. steamboat
		F. mechanical reaper
		G. air brake
		H. telephone

2. Be sure that each response is a plausible alternative for each premise.

A response may represent the same type of content as a premise, but still be implausible. For example, to students in an advanced science or history class it would be immediately obvious that Galileo did not invent the incandescent bulb. Were "Galileo" identified in the premise column and "incandescent bulb" in the response column, most students would eliminate incandescent bulb as implausible, even if they did not know Galileo's specific contributions or who invented the incandescent bulb.

3. Include more responses than premises.

This reduces the possibility of arriving at a correct answer by a process of elimination.

4. Keep the list of entries in each column brief.

Five to eight premises generally is considered an appropriate number. Response columns usually contain a couple more entries than this, as indicated in #3. above.

5. If the length of entries in the two columns is to be different, place the list with the longer phrases in the premises column.

Since the left hand premises column is read first, the shorter responses in the right column can be surveyed for the correct match. (See the example under guideline 8.)

6. When possible, arrange the list of responses in a logical order, e.g., alphabetical, numerical, or chronological order.

This makes it easier to locate a desired response, perticularly when many response options are provided.

7. Check that each premise has only one correct response (unless the directions warn students accordingly).

8. Write clear and complete directions.

Directions should indicate the basis on which the match is to be made, and whether a response can be used more than once.

Example:

Directions: For each of the definitions in column 1, place a letter for a sailing term from column II. A sailing term may be used only once or not at all.

Column I
Definitions

(E) 1. boat carrying mail and passengers on a regular schedule
(C) 2. crane used to lower lifeboats
(I) 3. fiber of flax or hemp
(H) 4. handle used to turn the rudder
(A) 5. heavy spar projecting ahead of the ship
(G) 6. sailboat having one mast
(D) 7. top of a ship's mast

Column II
Sailing Terms

A. bowsprit
B. brig
C. davit
D. masthead
E. packet
F. rigging
G. sloop
H. tiller
I. tow

(Hopkins and Antes, 1979, p.124)

A **checklist** for reviewing important aspects of matching items is presented in Table 9-3.

Fill-in-the-Blank Items

Fill-in-the-blank items are most commonly used to measure students' ability to remember content. These items are answered quickly and scored quickly. Writing fill-in-the-blank items is not always easy, however, for a sentence must be written in such a way that the word or phrase students are expected to supply can be easily inferred. Here are five guidelines for constructing fill-in-the-blank items.

TABLE 9-3

A Checklist for Preparing Matching Items

_____ (1) All entries within a column refer to the same type of content, e.g., causes or effects, events or dates, symptoms or ailments.

_____ (2) Each response is a plausible alternative for each premise.

_____ (3) There are more responses than premises.

_____ (4) The list of entries in each column is brief (about 5 to 8 premises, and a somewhat larger number of responses).

_____ (5) The premise column contains the entries with the longer phrases.

_____ (6) The list of responses is arranged logically.

_____ (7) Each premise has only one correct response (or the directions warn students accordingly).

_____ (8) The directions indicate the basis on which the match is to be made and whether a response can be used more than once.

1. **Be sure that only key words are omitted.**

Poor Item	Improved Item
The specific gravity of a (substance) is defined as the ratio of the density of a substance to the density of water.	The specific gravity of a substance is defined as the ratio of the density of a substance to the density of (water).

2. **Be sure that there is only one word or phrase that correctly completes the sentence.**

Poor Item	Improved Item
Chad is located in (Africa).	Chad is in the continent of (Africa).

3. **Limit the number of blanks per statement to one or two.**

If there are several blanks in a statement, the task is much more difficult and usually lacks clarity.

Poor Item	Improved Item
In (France), students in grade (six), take a national test to determine if they qualify for admission to a school called the lycée.	In France, students in grade six take a national test to determine if they qualify for admission to a school called the (lycée).

4. **Place the blank(s) near the end of the statement.**

If the blank(s) are placed at the beginning of a statement, students must read the sentence first to identify the task

5. **Avoid giving clues to the correct answer in the item.**

Check that the length of the blanks do not vary according to the length of the correct word or phrase. Also, check if the indefinite article "a" or "an" is needed before a blank. The article should be written in the form "a(n)" so not to indicate whether the correct word or phrase begins with a vowel or consonant.

A **checklist** for reviewing important aspects of fill-in-the-blank items is presented in Table 9-4.

TABLE 9-4

A Checklist for Preparing Fill-In-The-Blank Items

____ (1) Only key words are omitted; trivial words are not those left blank.

____ (2) There is one word or phrase that correctly completes the sentence.

____ (3) There are only one or two blanks per statement.

____ (4) The blanks are near the end of the statement.

____ (5) No clues to the correct answer are given in the item

- the length of the blanks do not vary according to the length of the correct word or phrase.
- if the indefinite article "a" or "an" is needed before a blank, the article is written in the form "a(n)" so as not to indicate whether the correct word or phrase begins with a vowel or consonant.

Short Answer or Exercise Items

Short answer items (exercises) allow students more freedom to respond to test questions or statements than either fill-in-the-blank or identification items. They may require students to write a sentence or two, draw a diagram, or solve a problem.

The main limitation of this item format is that it is inappropriate for measuring open-ended learning outcomes, i.e., those which call for divergent student responses. There should be a narrow range of acceptable variations in response to short answers/exercises. If teachers wish to assess students' ability to create original and extended responses to topics or problems, essays and product-related tests rather than short answers or exercises should be used.

Here are two guidelines for developing short answer items or exercises.

1. **Limit the number of correct responses to the item.**

Poor Item	Improved Item
Why did the United States invade Grenada?	List three reasons given by President Reagan for the U.S. invasion of Grenada.

Applying this guideline may involve indicating the precision or extent of the desired response, as suggested in the example below:

Poor Item	Improved Item
. What is the population of Tokyo, Japan?	. What is the population of Tokyo, Japan (round your answer to the nearest 100,000)?

2. **When appropriate, indicate to students the basis for awarding credit.**

Example:

"List the three principal sources of income for the Florida State Treasury." A student may receive one point for each source of income identified.

A checklist based on the two guidelines above is presented in Table 9-5.

TABLE 9-5

A Checklist for Preparing Short Answers or Exercises

____ (1) The number of correct responses to the item is limited.

_____ (2) When appropriate, the basis for awarding credit is indicated.

Essays and Product-Related Tests

Essays and product-related tests are useful in assessing students' ability to integrate and apply knowledge and skill in a meaningful way. In this section, three general guidelines for constructing these types of tests are provided.

1. Make sure that the topic, question or task has a definite focus.

Although essays and product-related tests vary according to the degree of freedom permitted students in composing a response, no test should allow so much freedom that students cannot determine where to begin developing their response. The following essay topic, for example, is so broad and vague as to make a focused response to it difficult: "Write an essay about the changing roles of women in society." If this topic were given as an inclass examination, students would need a large portion of the period simply to impose limits on the topic so that it could be addressed effectively. The scope and direction of the desired response simply is not apparent. Students would not know what the teacher was trying to assess through the question.

An example of an essay topic and a product-related task that provide a clearer focus are presented in Table 9-6.

Essays and product-related tests, of course, should not be so narrowly focused that students have no leeway in composing a response. If response limits are highly restrictive the task will amount to little more than a series of short answers or exercises.

2. Indicate limits on time, length, and sources of information.

Time is a particularly important factor on essay and product-related tests. Some teachers allow students to take these tests home, or they give test items to the students in advance to supplement the amount of time available in class. Whatever time constraints are to be placed on a test should be clearly communicated to students.

Table 9-6

An Illustrative Essay and Product-Related Task

Example 1: An Essay Topic

Should the graduation competence testing program used in this school be continued or terminated? Defend your position in two to three pages. Cite specific evidence about the positive or negative effects of the program on at least two of the following groups: students, parents, teachers, administrators, or prospective employers.

Example 2: A Product-Related Task

Design a plan for a new bike path in the city. The plan should reflect the following considerations:

- The voters of the city approved funds for the path by a narrow margin; the vast majority of senior citizens voted against expenditure of funds for this purpose.
- The path should not in any way interfere with automobile traffic downtown or around either of the two shopping malls in the city.
- The path should be well lit.
- The cost of developing the path should not exceed $10,000.

In presenting your plan, include a map of the proposed path that shows its position in relation to other bike paths, roads, and thoroughfares. Provide a rationale for your plan and a budget that lays out costs for planning and management, construction, and quality control.

Length is an important factor on essay tests, even though there obviously is no correct length for an essay. Mature writers recognize that length is influenced by the nature of the topic, the nature of the writer's personal style, the time and information available to develop a response, and so forth. Whenever feasible, however, providing students with a general idea of the expected length of an essay may help them organize their response.

Teachers also need to indicate the degree to which students may use notes, textbooks, and other sources of information to complete the test. The use of such material may be particularly appropriate when general essay writing or product design skill is stressed over knowledge of specific content.

3. Indicate how the test will be scored.

To the extent feasible the specific aspects of an essay or product to be evaluated, and the weight each aspect is to receive, should be indicated to students. For example, on an essay test students might be told that a maximum of ten points is to be awarded: 2 points for the clarity of the thesis presented, 3 points for the quality of the evidence used to support the thesis (sufficiency, relevance, and trustworthiness), 3 points for the logical organization of the argument, and 2 points for the precision and economy of wording.

Presumably, standards pertaining to each of these features of an essay would be discussed, demonstrated, and used in class before being applied for evaluation purposes. Students probably should not be evaluated on the "logical organization" of their essays, for example, unless specific elements of writing that make for logical organization had been studied previously. Studying examples of essays written with differing levels of logical organization would be a part of this learning process.

Observation of Performance

As an approach to testing, observation of performance involves recording and evaluating student performance as it occurs. While all tests provide information on performance, observations provide information on **in-process** performance, i.e., performance as it occurs.

The same general guidelines discussed in reference to essays and product-related tests apply to in-process performance evaluation procedures. The tasks students are asked to perform must have a **definite focus** and be accompanied by **appropriate limits.**

The task "Give an eloquent speech on a basic philosophic question," for example, probably would puzzle students rather than effectively guide their performance. Information about **relevant materials** also needs to be provided, e.g., a student would need a bat to complete an assessment of his or her batting technique. Limits on the **desired length** of student performance (a five-minute vs. a ten-minute speech in front of the class), and on the **information** to be used (a student may be instructed to use examples from his or her own life), are equally as important in evaluating in-process performance as in evaluating essays and other products.

Finally, students need to be informed about the procedures to be used in **assigning scores** to the performance. Many students may be under the impression that scores based on observations are highly subjective, if not arbitrary. Care needs to be taken to show students that clear and definable criteria guide performance evaluation.

Though product evaluations and in-process performance evaluations are similar in many ways, they differ in two important respects. One is that the **timing** of observations is critical for in-process performance and evaluations. If a behavior is missed, a teacher may not be able to observe it again during the time period allocated for the assessment. For example, a teacher may wish to observe students' actions in using a microscope in a biological laboratory. If the teacher fails to conduct observations during the first few minutes of the lab, significant behaviors, like wiping a slide with lens paper, placing a drop or two of a culture on the slide, placing the slide on the stage, turning to the objective of lowest power, and so forth, may go unrecorded.

Depending on what is being assessed, characteristics of student performance may vary substantially within a relatively brief period of time. This adds to the importance of timing as a critical factor in on-going performance evaluations. For example, a student may make three double faults during the first game of a tennis match, but make only one during the next eight games. Another student may be quite silent in the early stages of discussion, but become a frequent and active participant by the middle of the discussion. It therefore is important to specify the time period or periods in which observations are to occur, or to track behavior continuously throughout a performance evaluation.

Another distinctive characteristic of in-process performance evaluation is the potential problem posed by the **teacher's role as an observer.** Sometimes just observing a behavior causes students to behave differently than if they were not being observed. This can be a positive thing, as for example, when the teacher's presence during an achievement test stimulates students to work harder on a task than they ordinarily do. There are situations, however, in which the teacher's presence as an observer causes students to act in particularly artificial ways.

For example, a teacher may want to evaluate a student's performance during small group discussions. When the teacher is within earshot of the group, the student may appear tolerant of others' opinions and allow all group members a chance to speak. As soon as the teacher leaves the area, the student may engage in disruptive behaviors like ridiculing peers for their beliefs, or interrupting others while they are speaking. The teacher's presence thus would cause the student to behave in ways that contradicted his or her usual pattern of performance. Results from such an evaluation would be misleading.

When teachers suspect their observations may cause students to behave in a contrived fashion they may wish to ask students to evaluate each other's performance, or to record students' performance on film or tape. In the example cited in the previous paragraph, each student in the group could be given a form for counting, checking, or rating certain behaviors exhibited by other group members.

More specific information on preparing instruments for recording and evaluating observations, and for structuring the context in which observations are to be carried out, is provided in the next chapter.

REFERENCES

Chapter 9

Bloom, B. S., Madaus, G. F. & Hastings, J. T. Evaluation to Improve Learning. New York: McGraw-Hill, 1981.

Brown, F. G. Measuring classroom achievement. New York: Holt, Rinehart, and Winston, 1981.

Hambleton, R. K. & Eignor, D. R. A practitioner's guide to criterion-referenced test development, validation, and test score usage. (Laboratory of Psychometric and Evaluative Research Report No. 70). Amherst, Mass.: School of Education, University of Massachusetts, 1978.

Hopkins, C. D. & Antes, R. L. Classroom testing: Construction. Itasca, Ill.: F. E. Peacock, 1979.

TenBrink, T. D. Evaluation: A practical guide for teachers. New York: McGraw-Hill, 1974.

RELATED RESOURCES

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Gronlund, N. E. Measurement and evaluation in teaching. (4th ed.). New York: Macmillan, 1981.

Mehrens, W. A. & Lehmann, I. J. Measurement and evaluation in education and psychology, 3rd ed. New York: Holt, Rinehart and Winston 1984.

Priestly, M. Performance assessment in education and training: Alternative techniques. Englewood Cliffs, N.J.: Educational Technology Publications, 1982.

SBB9

CHAPTER 10

PREPARING, ADMINISTERING, AND SCORING TESTS

Introduction

The quality and utility of information from a test depends not only on the characteristics of individual items or procedures, but on the design of the test as a whole; the conditions under which the test is administered; and the procedures through which it is scored. This chapter provides guidelines for test preparation, administration, and scoring.

The chapter is the longest in the handbook. To aid the reader in identifying sections of particular interest, an outline of the guidelines and topics addressed is presented below.

I. Guidelines for Preparing Tests (pages 105 to 108)

  A. Clarify the outcomes to be assessed and the emphasis each is to receive on the test.

  B. Determine an appropriate test length.

  C. Assemble test items in an appropriate form and sequence.

  D. Make sure that test directions are clear and complete.

II. Guidelines for Administering Tests (pages 108 to 112)

  A. Consider the purpose to be served by a test when developing procedures for test administration.

  B. Consider special issues that may arise when administering essay and product-related tests or in-process performance evaluations.

III. Guidelines for Scoring Tests (pages 112 to 124)

  A. Use scoring procedures that respond to particular information needs.

  B. When appropriate, develop answer keys to score response-completion items.

  C. Develop special procedures for evaluating essays, products, and in-process performance evaluations.

    1. checklists
    2. rating scales
    3. frequency counts

D. Use scoring procedures in a careful and consistent manner.

1. when evaluating essays and products
2. when evaluating in-process performance

Guidelines for Preparing Tests

Previous chapters have focused on the implications of different types of learning outcomes for test design, and on standards of technical adequacy that test items should meet. The guidelines in this section are intended to assist teachers in selecting and organizing items for use in a test as a whole. The guidelines draw attention to a number of factors that need to be examined when building a test, such as the range of outcomes to be assessed on a test and the emphasis each is to receive, and the length and layout of the test.

Clarify the Outcomes to be Assessed and the Emphasis Each is to Receive on the Test

The first step in preparing a test is to identify the particular outcomes to be assessed. This may be a rather straightforward matter, as, for example, when one wishes to assess what students have learned from a single lesson or a small number of lessons. However, determining the outcomes to be assessed on tests covering broad segments of instruction, for example, end-of-semester or end-of-year tests, can call for considerable judgment, because in such cases there is a wide range of potential outcomes that might be assessed. On tests with a broad focus, a teacher may wish to assess general outcomes that call for the integration of many smaller outcomes, or to select a representative sample of both general and specific outcomes. Whatever approach a teacher uses, it obviously is crucial that students be informed about the outcomes to be assessed, and the logic underlying their selection, well in advance of test administration.

In addition to determining which outcomes are to be assessed, one needs to establish the relative emphasis each outcome is to receive on the test. A teacher may, for example, wish to give greater emphasis to broader, long-term outcomes than narrow outcomes.

The relative emphasis of an outcome can be indicated on a test in one or two ways, or a combination of each. One is to vary the **number of items** for an outcome according to its relative importance. For example, on a particular end-of-semester test, all highly important outcomes might be assessed with five or more items; all outcomes of moderate importance might be assessed with two to four items; and outcomes considered least important might be measured with a single item or no item at all. However, it also is possible, and often desirable, to assign different **weights or point values** to items depending on the importance of the outcomes they are intended to assess. The value of this practice is especially evident when using essays and product-related tests in combination with objective items. A single essay item may be used to assess one broad, particularly important outcome. Such an essay obviously should carry more weight in calculating a student's score on a test than an individual true-false or fill-in-the-blank item. The point values assigned items or procedures on a test need to reflect the relative importance of the learning outcome on which they are based.

Determine an Appropriate Length for the Test

The length of a test depends on (a) the uses to be made of test findings; (b) the nature of the outcomes to be assessed; (c) the characteristics of the students who are to take the tests; and (d) the time available for testing.

The length of a test is fundamentally influenced by the anticipated uses of test results. In general, tests that are to be used to make important judgments about a student, for example, judgments about a final grade, are longer than tests that have less serious or long-term implications for a student. This is partly because tests carrying important implications usually cover a relatively broad band of learning outcomes. However, another reason for the increased length of tests used to make major decisions is that, all other things being equal, longer tests provide a more reliable, i.e., trustworthy, indication of what and how well students have learned than shorter tests. As a simple illustration, an eight or ten-item test of a student's skill in solving mathematical problems involving the concepts of time, distance, and rate would produce more reliable information on the degree of the student's skill mastery than a two or three-item test. Similarly, a 70 item final examination over the goals for a course as a whole probably would produce a truer picture of what students have learned in the course than a 40 item test over these same goals.

The nature of the outcomes to be assessed also influences the length of a test. Highly specific outcomes often can be assessed with a very small number of items. For example, consider this outcome: "Students will be able to define the concepts of alliteration and assonance." One or two items would be sufficient to provide reliable information on whether or not students know these definitions. Broader outcomes, however, usually require a larger number of items, or a more extensive or in-depth sampling of student's behavior or work.

The characteristics of students and the time available for testing also have direct implications for test length. Students with limited attention spans and poor powers of concentration obviously cannot be expected to respond effectively to long tests, e.g., tests that are two or three hours in length. Generally, the more mature the learners the greater will be their capacity to take long tests.

Finally, it is recognized that the length of a particular test often is determined by school schedules and policies.

Assemble Test Items in an Appropriate Form and Sequence

Test items need to be arranged and presented in a deliberate fashion. Here are some general suggestions that may aid in this process.

1. The correct response to one item should not be given or implied in another item.
2. The medium you have chosen for item presentation should be appropriate. (Print is the most widely used medium for testing, but in some cases it may be more meaningful or efficient to present items orally, write them on the board, project them on a screen, or present them by computer.)
3. Items should be placed in a logical order on the test, e.g., grouping items by learning goal or objective or by item format.
4. No item should be split across two pages.
5. Introductory or reference material for an item should be placed on the same page as the item, or on a facing page.
6. Items should be spaced so that they can be read, answered, and scored with minimal difficulty.

Make Sure that Test Directions are Clear and Complete

Here are some guidelines to follow when preparing test directions:

1. Directions should indicate where and how students are to record their answers, e.g., "Mark all your answers on the separate answer sheet provided. Mark through the answer you think is correct in this way:"

   Example: A ~~B~~ C D

2. Directions should be sufficiently complete and precise so that students will know the:

   - time limits
   - resources (books, notes, etc.) permitted
   - points each item is worth
   - whether or not to guess
   - special procedures that are to be followed, e.g., students are to show their work for a problem in addition to providing a solution

3. If students are unfamiliar with a particular type of item, a sample of the item type should be included.

II. Guidelines for Administering Tests

In this section, guidelines for establishing appropriate test-taking conditions are provided. The guidelines indicate that the uses to be made of test results influence the way in which tests are administered. They also suggest that complex issues often are involved in using product-related tests and in-process performance evaluations.

Consider the Purpose to be Served by a Test When Developing Procedures for Test Administration

The uses to be made of test findings have implications for the context in which tests are administered. When a test is to be used, for example, to certify that a student has attained an important set of outcomes, or to assign grades to students it is essential that each student takes the test under the same set of conditions. If one student is allowed as much time as he or she needs on a semester exam, for instance, while another student's time is restricted, differences in these students' test scores may be due as much to test taking conditions as student learning. Guidelines for administering objective tests in formal testing situations are presented in Table 10-1.

Informal assessments, like quizzes or homework assignments, however, do not require such standardized conditions, since the reliability or comparability of the scores obtained from informal measures is less critical. If differences in administrative conditions, e.g., one classroom is warmer and more comfortable than another, lead to some differences in performance on informal tests, the consequences for students are small. In fact, for assessment procedures like homework assignments, it would not be desirable to prescribe too tightly the way in which tasks are to be carried out. Some

TABLE 10-1

Guidelines for Administering Objective Tests In Formal Testing Situtations

1. Check to see that the classroom is comfortable, i.e., conducive temperature, lighting, ventilation.

2. Be sure that all talking stops at least a minute before the test is distributed, or the task is presented (unless, of course, the test involves oral assessment).

3 Take steps to minimize interference and disruptions from outside the classroom, e.g., post a "Testing, Do Not Disturb" sign on the door, and tell students in advance the procedures they should follow if they are late for class the day of the test.

4. Have a visible supply of pencils, pens, paper, and other test-related material available, or be sure that students have brought with them all they need to complete the test.

5. If the test is timed, be sure that there is a visible and accurate clock in the room, or that you regularly post the time on the board.

6. Unless the test is to be open-book, ask students to remove non-testing material from their desks.

7. Pass out the test or present the task in a systematic and efficient fashion.

8. Be sure students know what they are to do if they complete the test before the time is up.

9. Be sure students understand the uses to be made of test results.

10. Collect test papers or other student products in a systematic and efficient fashion.

11. If the test is being administered in more than one location or at more than one time, be sure that the time permitted, the directions and assistance provided, and other testing conditions are as similar as possible in all test administrations.

12. Establish clear make-up procedures for students who miss a test, or identify the conditions under which students need not make up a test or quiz.

variation in the conditions under which individuals respond to informal assessment tasks may be unavoidable, and in some instances may be desirable.

Consider Special Issues that may Arise when Administering Essay and Product-related Tests, and In-Process Performance Evaluations

There are three issues that may be encountered when administering essay and product-related tests, and in in-process performance evaluations, that rarely are encountered when administering objective tests. These involve: (1) having to deal with the presentation of complex topics; (2) use of scarce materials or equipment; and (3) the use of group projects.

1. Complex Topics or Situations

In the great majority of cases, tasks prescribed on objective tests are relatively straightforward and stated in writing. Beyond the guidelines listed in Table 10-1, administering tests of this kind usually involves little more than passing out a test booklet or other printed material.

Product-related tests, however, and in-process performance evaluations frequently involve topics and situations that are presented in a non-print medium. This is particularly true in fine arts and industrial arts, where a test stimulus may take the form of a live model, a malfunctioning television set or an evocative piece of music.

When topics or situations are presented in such complex forms special care needs to be taken to assure that the critical aspects of the test stimulus are communicated to all students, and that the stimulus is presented for an appropriate period of time. For example, an art teacher may wish students to sketch the profile of the person sitting next to them in class, but some student-models may sit still while others may move abruptly or repeatedly. When this occurs the stimulus conditions vary from one student to the next, and it becomes difficult to interpret results from the sketching exercise. The quality of a sketch might depend as much on the quality of the peer-model as on a student's artistic skill.

When feasible, teachers may wish to tape or film situations that otherwise are not readily standardized. If human action is to be a part of the test situation, rehearsals may be appropriate to assure that people perform in the fashion desired. Scripts that prescribe statements, responses, and movements may also be beneficial.

2. Use of Scarce Materials or Equipment

Just as the stimulus materials for product-related tests and in process performance evaluations are likely to be complex, so too are the resources needed to create an effective response. For example, on a test over a unit in pottery, students may need access to a potter's wheel, a kiln, and several pounds of clay. A unit test on sewing in a home economics class may require tape measures, patterns, yards of material, a variety of threads, and a large supply of sewing machines.

When resources to be used for assessment purposes are in limited supply, students' access to the resources becomes an issue. A schedule for using materials or equipment is needed to assure that each student has an equal opportunity to work with them. For example, assessment may need to be stretched out over several days to give each student access to a particular piece of equipment. In addition, the effects of students observing each other use test-related resources needs to be considered. For example, the last student to use a potter's wheel may produce a higher quality piece of pottery than the first person simply because he or she has the benefit of watching others make and correct mistakes on the wheel. Use of scarce resources should be scheduled and sequenced in such a way that no students have an advantage over others because of when they have access to needed resources.

This same general point also applies in the context of take home exams. Students typically have different levels of access to resources outside of class. One student may live next door to a library, another may have parents who are highly knowledgeable in the area being tested, while still another may have a well-equipped studio, shop, darkroom, etc., in which to create products. Either the outside resources students are to use need to be limited to those commonly available, or a topic or problem should be assigned that does not call for specialized resources that may be distributed unequally among students.

3. Administration of Group Projects

On some product-related tests, students may be asked to work as a team. For example, students may be asked to create a topographical map, design a playground, or paint the walls of the school cafeteria as a collective effort. This may promote cooperation, teamsmanship, and a feeling of community among students. On the other hand, it may lead to some students standing around with nothing to do, other students catching up on assignments from another class, or one or two students doing most of the work called for in the project. If group projects are to be assigned and assessed, it generally is desirable to "equate" the groups, so that each represents a variety of student abilities and talents. It also may be desirable to specify different roles for students to assume in the project, e.g., the role of manager, craftsperson, technicial editor, etc., and to describe the tasks and timelines associated with each role. Mature or experienced students might be expected to do this level of planning on their own, but even for these students some indication of how students should work together is helpful.

Guidelines that structure students' involvement in a group project usually are built into the test or assignment itself, that is, teachers prepare guidelines prior to administering the test. Even so, teachers should check carefully whether students are following the guidelines. Product-related tests to be completed as a team effort usually require the monitoring of students' behavior more actively than do objective tests or other types of tests requiring individual work.

III. Guidelines for Scoring Tests

Rules for assigning scores to student performance are a critical aspect of a test. The nature of these rules depends on a number of factors, including the uses to be made of test information and the format of the test. Guidelines for developing test scoring procedures are offered below.

Use Scoring Procedures that are Appropriate for Particular Information Needs

The type of score that one assigns a test paper, product, or performance should reflect the uses to be made of test scores. The more detailed information one needs about student achievement, the more specific are the scores that need to be produced. Consider, for example, the difference between a test designed to determine a student's strengths and weaknesses in a particular learning area and a test designed to determine whether or not a student should get credit for a course. In the first case, it may be helpful to calculate a score for each outcome or type of outcome assessed on the test. Reporting scores for each kind of learning outcome might help both the teacher and the student pinpoint specific areas of strength and weakness. Examples of test scores that are anchored to specific types or levels of learning outcomes are shown in Table 10-2.

Specific scores of the kind illustrated in Table 10-2 would have little utility, however, if the purpose of a test was to determine, for example, whether or not a student should get credit for a course. In this case, it is the students' overall performance on a test that typically is of primary concern. Information about performance on one set of items versus another is not generally relevant to the decision as to whether credit should be granted. A single score that reflects performance on the test as a whole may be specific enough to guide the decision at hand. Thus, the kind of scores to be assigned on a test should be based on teachers' particular information needs.

When Appropriate, Develop Answer Keys to Score Response-Completion Items

As indicated in Chapter 5, scoring fill-in-the-blank and short-answer items involves only a small amount of subjective judgment, and these items therefore have been classified as objective items in this handbook. However, on many short-answer items students may a) express the correct answer in different ways, b) get part of the answer right and another part wrong, or c) omit part of the answer. For example, a test item may ask students to list four steps involved in photosynthesis. A student could list two of the steps, but leave others out. Guidelines for awarding full or partial credit need to be indicated on an answer key and, to the extent feasible, communicated to students in the test directions.

One way to assure that the answer appearing on an answer key is as good as it can be is to check the key against the answers provided in a sample of test papers. Students may generate good answers that teachers do not anticipate, or their answers may indicate the need to refine the standard used to award full or partial credit. Of course, all papers should be scored after an answer key has been revised.

Develop Special Procedures for Scoring Essays, Products, and In-Process Performance Evaluations

Procedures need to be developed to reduce the subjectivity involved in evaluating student responses to essays, product-related tests, and in-process performance evaluations. The most common of these procedures are **checklists** and **rating scales.** **Frequency counts** also are used when teachers want to determine the number of times a particular behavior occurs in a specified time period.

Checklists. Checklists are used to indicate whether desired aspects of a product or a performance are present or absent, accurate or inaccurate, appropriate or inappropriate, and so forth. Here are some guidelines for building such checklists.

1. **Identify the features of the product or performance to be assessed.** These may be highly specific qualities or behaviors, or more general ones. As an illustration of a highly specific list of behaviors, a portion of a well known checklist for assessing skill in using a microscope (Tyler, 1930) is printed in Table 10-3.

TABLE 10-3

An Illustrative Checklist for Assessing Skill in Using a Microscope

	Student's actions	Sequence of actions
a.	Takes slide	1
b.	Wipes slide with lens paper	2
c.	Wipes slide with cloth	
d.	Wipes slide with finger	
e.	Moves bottle of culture along the table	
f.	Places drop or two of culture on slide	3
g.	Adds more culture	
h.	Adds few drops of water	
i.	Hunts for cover glasses	4
j.	Wipes cover glass with lens paper	5
k.	Wipes cover with cloth	
l.	wipes cover with finger	
m.	Adjusts cover with finger	
n.	Wipes off surplus liquid	
o.	Places slide on stage	6
p.	Looks through eyepiece with right eye	
q.	Looks through eyepiece with left eye	7
r.	Turns to objective of lowest power	8

Detailed scoring instruments like this usually are used only in tests that set tight limits on students' responses. The checklist for assessing skill in using a microscope, for example, probably would not be used unless students were requested to follow the step-by-step procedure outlined on the scoring instrument. Were students asked merely to "use the microscope in a systematic and responsible manner," there would be little justification for evaluating their performance on the specific behaviors identified on the checklist. Sharply focused checklists are appropriate only for assessing performance on sharply focused tasks.

Here is an illustration of a checklist with more general descriptions of desired traits. It is intended to be used as a tool for evaluating a meal planned by a student in a home economics class.

	The meal is:	**YES**	**NO**	**Comments:**
1.	nutritious	____	____	____________ ____________
2.	flavorful	____	____	____________ ____________
3.	low in calories	____	____	____________ ____________
4.	easy to prepare	____	____	____________ ____________

Presumably this checklist would not be used to assess student learning during a unit or course unless the class had discussed in advance the meaning of such broad tems as "nutritious" and "flavorful."

Teachers may wish to include on a checklist undesired features of a performance or product, as well as desired features, if these are likely to appear in a student's work and seriously detract from its quality.

2. List the features to be evaluated in a meaningful order. There is no one correct order for items in a checklist to be organized, but some rule of organization should be followed. This could be by form or content, by importance, or by the order in which the evaluator will see the features. For example, a checklist designed to assess an advertisement for a commercial product might focus first on whether or not the name of the product was prominently displayed, then on whether the purposes to be served by the product were made clear, and finally on whether the benefits accompanying the product were shown to be greater than benefits coming from competing products.

3. Identify the two options for responding to the features listed. The most common options on checklists are yes-no, present-absent, and correct-incorrect. Other response options, however, certainly are possible, e.g., adequate-inadequate, valid-invalid, objective-biased, neat-sloppy.

4. Leave enough space for comments. Since checklists permit only black-and-white responses, some of the scorer's observations and insights may not be reflected on the instrument. A space for comments, either next to each feature or at the bottom of the checklist, provides at least a partial remedy for this limitation.

Rating Scales

Using rating scales to evaluate a product or performance allows distinctions to be drawn among degrees of quality represented. To many teachers, this is a distinct advantage to the all-or-none quality of a checklist. Guidelines to be followed in constructing rating scales are presented in the next several pages.*

1. Identify the aspects of a performance or product that are to be rated. As in the case of checklists, or any other approach to evaluation, the first step always is to identify what is to be evaluated. Here are two examples that apply to the use of rating scales.

Example 1 An English teacher wants to rate the quality of ideas, organization, wording, and style used in an essay.

Example 2 A physical education teacher wants to rate the grip, stance, stride, and arm movement used in swinging a baseball bat.

2. Decide how much detail is to be used to describe the points on a scale. Some rating scales provide detailed descriptions of the qualities associated with each point on the scale. Others provide highly general descriptions, or simply use numbers to represent different levels of quality. Two different scales for rating the use of language in a short story are presented below.

---

* These guidelines are adapted from Priestley (1982), and TenBrink (1974).

Use of Language in a Short Story

Example 1

1	2	3
poor	average	good

Example 2

1	2	3
dull wording and monotonous sentence structure	occasionally vivid wording and some variety in sentence structure	vivid wording and wide variety in sentence structure

Scales like the one in example 2 provide more useful feedback to students than the scale used in example 1. In addition, ratings are more apt to be reliable when scale values are associated with specific descriptions. As a rule, it is desirable to anchor the points on a rating scale with as much specificity as possible.

The degree of detail appropriate on a particular scale, however, depends in part on the nature of instruction students have received regarding the processes or products to be assessed. For example, a class may have studied at great length various patterns of language used in short stories. Standards for appraising the appropriateness, imaginativeness, effectiveness, or elegance of language may have been made quite explicit through the instructional process. In this kind of situation, a scale with little descriptive detail may be adequate since students would know from past experience what a score of "3," "2," or "1" on a language usage scale signifies. In most cases, however, it would be best to carefully spell out evaluative criteria on the rating scale itself, if only to confirm the appropriateness of the criteria students have studied and applied in the past.

In addition to specifying rating scale values in writing, a teacher may wish to construct or select illustrative papers or products against which to rate a student's work. Essays, for example, are frequently rated by comparing them with "standard" essays that represent fixed points on a rating scale. Illustrative products serve to anchor points on a rating scale.

3. **Decide how many points to include on the scale.** The number of points to include depends on the subtlety or fineness of the distinctions in quality one wishes to draw. In some cases, a scale with a few number of points, e.g., three, may be appropriate (see Example 2 on the top of this page), whereas in other cases, a scale with a larger number of points, e.g., five or six, is called for (see the "Improved Scale" on the next page).

4. Define the extremes of the scale. The end points on a scale sometimes express frequencies, e.g., "0-2 typing errors," "over 10 typing errors." When this is the case, end points are relatively easy to define.

When end points are defined in qualitative terms, however, the task of describing end points is more demanding. This is illustrated by the following example (adapted from TenBrink, 1974, p. 282).

A Salt and Flour Map

1	4
Severely cracked; poorly proportioned; inaccurate borders	Smooth; well proportioned; highly accurate borders

5. Describe the options between the extremes. The difference in values between any two points on the scale should be about the same. For example, the points on the "poor" scale below represent unequal units; the differences in achievement represented by points 3 and 4 are much larger than differences in achievement represented by points 1, 2, and 3:

Poor Scale

1	2	3	4
Unreadable	Sloppy but readable	Readable	Elegant

Improved Scale

1	2	3	4	5
Unreadable	Sloppy but readable	Readable	Readable but stylish	Elegant

(Priestley, 1982, p. 140)

6. Arrange the scale in a meaningful order. When more than one scale is to be used to evaluate a product or performance, the scales need to be arranged in a meaningful order. Scales most commonly are arranged in a row, with each scale having the same direction, e.g., negative to positive, weak to strong, acceptable to unacceptable. It is easier for a teacher to mark scales in a series if they are arranged in the same direction. If a "4" on one scale is a negative rating and a "4" on a scale just below is a positive rating, there is a potential for the scorer to become confused.

However, there is also a potential problem when scales follow the same direction. Scorers may develop a response set, i.e., a tendency to mark each scale about the same. For example, if a scorer rates features of a student's work as very high on several consecutive scales, he or she may fall into a pattern of assigning high ratings on subsequent scales, even if ratings are undeserved. To reduce the likelihood of a response set developing, the direction of the scales can be shifted every so often, e.g., from "negative to positive" and later from "positive to negative."

7. **Be clear about how the scale is to be marked, e.g., circling a number, checking a space, etc.**

8. **Leave enough space for comments, student's name or identification number, the date and for recording total points.**

Frequency Counts

In evaluating in-process performance, a teacher may wish to determine the number of times a particular behavior or set of behaviors occurs. This is called a "frequency count." Counting and recording the number of times a student paraphrases another student's comments during a discussion, or the number of chinups a student can do in a minute, are examples of frequency counts. Frequencies of undesirable behaviors or errors, e.g., the number of times a student strikes out in a softball game, also can be made.

Perhaps the most important thing to keep in mind when developing a procedure for counting frequencies, other than to specify clearly the behaviors to be counted, is to limit the number of behaviors to be observed in any one period of time. If too many actions are to be counted observations may become unreliable. They also are likely to produce frustration and irritation on the observer's part.

Use Scoring Procedures in a Careful and Consistent Manner

Checklists, rating scales, and procedures for obtaining frequency counts need to be used with the same degree of care that goes into their creation. Guidelines for using these instruments in evaluating (a) papers and products and (b) in-process performance are presented below.

Guidelines for using checklists and rating scales and frequency counts in evaluating papers and products

1. **Maintain student anonymity.** A teacher may ask students to place their names on a separate front cover of a paper to be scored . These cover pages can then

be quickly turned back before the paper is read. Or, more simply, teachers may make a deliberate effort to avoid looking at students' names while evaluating their work. Such efforts reduce the chances that the scorer will evaluate a paper or product on the basis of what he or she knows or feels about a student rather than on the basis of the quality of the student's work.

2. **If more than one product-related task is given to each student, evaluate all responses to the first task before evaluating students' responses to the second.** This minimizes the possibility of a response set developing, i.e., a scorer rating all products produced by a particular individual in the same way he or she rated the first product produced by that individual.

3. **Group papers/products into tentative score groups before assigning scores.** Before actually assigning scores it generally is a good idea to scan all papers/products to develop a general awareness of the range of quality represented. Rough grouping of papers or products can be formed on the basis of this initial judgment of quality. For checklists, only two groups need to be formed. For rating scales, groups need to be formed representing each point on the scale. As a rule, it is easier to form extreme groups first, i.e., papers/products that have the characteristics described by the end points of the scale, and then to identify those representing the points between the extremes.

When initial score groups have been established, the papers or products placed in each group should be reviewed carefully to see that they are at about the same level of quality. Groupings are then refined as needed. A score is assigned to each paper/product according to which group it finally is placed within.

It should be noted that this procedure is based on the assumption that papers or products will be produced that represent each of the points on a scale. This assumption may be invalid, however. If instruction is highly successful, all students in a class may produce work of quality, and only the highest one or two points on a scale may be assigned. Similarly, if instruction is unsuccessful, most students may produce poor work. and only the lowest point or two on a scale may pertain. To the extent that the qualities or behaviors assessed on a scale are clearly defined, a teacher can evaluate papers/products in relation to the scale rather than in relation to each other.

4. **When important decisions are to be made on the basis of test results, use more than one judge to assign scores.** If two people score a paper or product independently and arrive at the same score or a similar score, more confidence can be placed in the score than when only one person produces the score.

Generally, people who will be scoring as a team prepare by studying a sample of the products to be scored. Members of the team practice scoring the sample and discuss any discrepancies in scoring. In this way, a common understanding of the meaning of each response option on a list or scale is established.

When two scores are assigned a product or performance, the scores generally are summed. For example, if one person gives a mechanical drawing a score of "3," and one gives it a score of "4," the student would receive a "7" on the product. If the two scores assigned are unacceptably different, the scorers usually meet to discuss their differences and to arrive at an appropriate score. In these cases where products are concerned or third scorer may also be asked to evaluate the product and help resolve the initial difference in scores.

Guidelines for using checklists, rating scales, and frequency counts in evaluating performance

1. **Limit observations to the qualities or behaviors specified.** Students generally exhibit many more behaviors or traits during an in-process performance evaluation than are identified on a checklist or rating scale. For example, a rating scale designed to measure speaking techniques like eye contact with the audience, voice modulation, and use of hand gestures may make no reference to other features of performance, such as enunciation and rate of speaking that affect the overall quality of a presentation. Nonetheless, if the checklist or rating scale is to be used according to its intended purpose, observations should focus only on the features of performance explicitly indicated on the scoring instrument. It is better to carefully observe a few aspects of performance at a time than to observe many qualities or behaviors at once and risk obtaining inaccurate information. This is not meant to imply, however, that it is inappropriate for observers to record in the comments section a particularly noteworthy feature not identified on the recording form.

2. **Practice using observational procedures before applying them in a test situation.** When procedures for observing behavior have the potential for producing anxiety, self-consciousness, or other undesirable consequences for students, the teacher may wish to practice the procedure with colleageues until he or she can use it smoothly and efficiently. Teachers who attract a great deal of attention to themselves when placing checks, marks, or comments on a scoring instrument, or who move clumsily or conspicuously around the room to gain a vantage point on student performance, are not ready to use observation for testing purposes.

For their part, students may benefit from "trial runs" with procedures for assessing in-process performance. As they gain experience in being observed, students may focus less on the action being taken to record observations and more on the quality of their own performance.

3. **Make multiple observations whenever possible.** Observations are sometimes hit-and-miss. The observer often has only a brief period of time to record a particular set of behaviors or qualities. Within the time available for an observation, the behaviors or qualities of interest may not be exhibited, but this does not necessarily mean that the student lacks the skills desired. For example, a physical education teacher may have only three or four minutes on a given day to observe a student's skill in fielding line drives in softball. During these three or four minutes, the batter may hit only a few good line drives. Judging a student's skill on the basis of this small sample of performance would be inappropriate. A teacher needs to obtain many observations of performance before valid inferences can be made about a student's level of skill.

4. **As in the case of product evaluation, when important decisions are to be made on the basis of observations, use more than one judge to rate student performance.** The general procedures described earlier for evaluating papers or products in a team apply to the observation of in-process performance. However, instead of collecting concrete papers or products, evaluators of in-process performance may make a videotape of the behaviors observed so that raters have a basis for identifying and analyzing sources of discrepancy in scoring.

REFERENCES

Chapter 10

Priestley, M. Performance assessment in education and training: Alternative techniques. Englewood Cliffs, N.J.: Educational Technology Publications, 1982.

TenBrink, T.D. Evaluation: A practical guide for teachers. New York: McGraw-Hill, 1974.

Tyler, R.W. A test of skill in using a microscope. Educational Research Bulletin, 1930, 9, 493-496.

RELATED RESOURCES

In addition to the references and related resources listed at the end of Chapter 9, the reader may wish to consult:

Bock, D.G. & Bock, E.H. Evaluating classroom speaking. Urbana, Ill.: ERIC Clearinghouse on Reading and Communication Skills, 1981.

Educational Measurement (Special issue on writing assessment). 1984, 3 (1).

Najimy, N.C. (Ed.) Measure for measure: A guidebook for evaluating students' expository writing. Urbana, Ill.: National Council of Teachers of English, 1981.

Stiggins, R. Evaluating students by classroom observation: Watching students grow. Washington, D.C.: National Education Association, Professional Library, 1984.

SBB10

CHAPTER 11

USING TEST INFORMATION AS FEEDBACK TO STUDENTS

Introduction

One of the most important benefits to come from classroom tests is the feedback they provide students on what and how well they have learned, and on what learning gaps need to be filled. In this chapter, six different themes regarding the use of test results as feedback to students are discussed:

1. Using tests with built-in feedback procedures;
2. Using test scores as feedback;
3. Commenting on test performance to supplement and enlarge upon the reporting of test scores;
4. Using audio or video recordings of student performance as feedback;
5. Preparing students to provide feedback to each other; and
6. Providing feedback on test performance to the class as a whole.

The focus of the first five topics is on giving feedback to individual students on their test performance. The assumption uderlying the chapter is that personalized feedback is indispensable to a student's learning. Feedback on the performance of a class generally may supplement feedback on individual performance, but it cannot substitute for it.

Using Tests with Built-In Feedback Procedures

Some tests provide direct feedback to students. Tests accompanying computer-assisted instructional programs, for example, typically score student responses immediately and display scores in the form of lights, images, and sounds. Computer programs also have been developed in tightly structured subject areas, like the basic skills of mathematics, that present problems to students, identify error patterns, and prescribe corrective measures. Self-scoring tests that furnish information on the reasons one response to an item is superior to another also have been designed. When feedback procedures are integrally tied to testing procedures, tests themselves are important teaching tools.

Using Test Scores as Feedback

In this section, two points are made that build upon the discussion of test scores contained in Chapter 10: (1) scores on individual sections of a test (subtest scores) generally are more useful as aids to student learning and instructional decision-making than total test scores; and (2) scores used to indicate the quality of student products or performance that are anchored to specific qualities or behaviors are more useful than scores that are not so anchored.

Subtest scores usually have more value as feedback than total scores because they indicate how well or poorly students did in particular areas of learning or with respect to particular categories of learning outcomes. Scores identifying the kind of content or level of performance with which students had difficulty are potential aids to both teachers and students in planning improvement strategies. Similarly, scores that indicate students are highly knowledgeable in a specific content area, or are able to perform skillfully at a particular level of use, may assist teachers and students in deciding what new topics or types of learning to pursue.

When determining the appropriate number of subscores to report on a test, however, it must be recognized that the larger the number of subscores reported, the smaller is the sample of behavior on which each subscore is based. Since scores based on small samples of behavior, e.g., two or three items, are more apt to provide misinformation than scores based on large samples of behavior one must balance the benefits from subscores with the risk involved in reducing the sample of behavior from which these scores are derived.

With respect to ratings of student products or performance, ratings with anchored values are more illuminating to students than highly general ratings. A score of "2" on a 4-point rating scale used to assess the quality of a poem, for example, communicates effectively only if the features of a poem represented by a score of "2" are specified. Numbers in themselves are not particularly illuminating sources of feedback.

In many cases of product or performance evaluation the scoring keys, checklists, rating scales, or frequency counts that the teacher uses to assign scores may be presented to students so that they can see clearly the basis of a score. These scoring devices may be distributed and discussed when teachers are communicating to students the nature of the learning outcomes they are expected to achieve. For example, a writing teacher may not only inform students that they will be learning to write clear and coherent essays, but may pass out a rating scale that defines concretely different levels of clarity and coherence. To the degree that students discuss and apply this scale during the course of writing instruction and practice, they will be better prepared to interpret scores on writing exams based on the scale.

Commenting on Test Performance to Supplement or Enlarge Upon the Reporting of Test Scores

Even if test scores are anchored clearly to areas of content, levels of performance or desired qualities or behaviors, supplementary comments on test performance can be helpful to students. Comments may serve any or all of the following purposes:

- To highlight and reinforce the message conveyed by scores, e.g., "You did very well on the story problems, but you seemed to have difficulty with items on graphing relationships."
- To place scores in a broader and more personal context, e.g., "You did a lot better on this unit test than on the last quiz. You certainly have thought more thoroughly about the relation between abolitionism and liberal Protestantism. Rereading William Lloyd Garrison's address paid off!"
- Provide more detailed information about a student's test performance, e.g., "The metaphors you use in the second paragraph are particularly well-drawn."
- Suggest steps for improvement, e.g., "You missed most of the items on the origin of viruses in plants. I'd suggest you reread the handout on this topic and listen to the accompanying cassette. When you complete these steps, you'll be ready to take an oral quiz on plant viruses.

- Signal the need for a conference with the teacher or for some sort of diagnostic assessment, e.g., "You have scored so low on these last two units that I am wondering what's going wrong? Please see me at the end of class."
- Indicate to a student that he or she is ready for new or more independent work, e.g., "Now that you've got this under your belt, you're ready to move on to the next unit."

Using Audio or Video Records of Student Performance

It may be feasible to record students' behavior during some performance evaluations. This allows the teacher and the student, or the student independently, to go over the performance in private at a later date. The tape can be stopped at various moments that illustrate a particular strength or limitation in the student's performance. The student can see or hear the specific evidence upon which the teacher based the score.

It may be impractical to record every student's performance or unrealistic to expect every student to have time and opportunity to review a tape after class. If tapes are to be used as a regular feature of instruction and assessment, perhaps a different sample of students can be selected for each taping session.

Taping may distract some students or cause them to be unduly self-conscious. However, research on student responses to classroom videotaping indicate that this is not a common reaction (Bush, Bittner, & Brooks, 1972). Further, students appear to gain a great deal from analyzing a record of their own performance.

Preparing Students to Provide Feedback to Each Other

Students may be trained to score and provide feedback on interpretive tests and performance evaluations. Learning to evaluate and discuss characteristics of products and performance can help students internalize desired standards of quality, accuracy, speed, etc. Also, when students receive feedback from a peer that is consistent with the kind of feedback supplied by the teacher, they are more likely to recognize that scoring is not a matter of mere personal preference or taste, but of the consistent application of criteria.

However, evidence indicates that without training, students by and large do not provide effective instruction or feedback to peers (Niedermeyer, 1976). Evidently, untrained students often fail to maintain positive rapport with their peers or to correct or to deal appropriately with correct or incorrect responses.

The most successful training programs for peer tutors (which have involved training in instructional methods generally, and not simply in giving feedback) have involved structured role playing. During role playing, it is common for students to tutor each other using scripted material. The script illustrates appropriate responses a tutor might make to various situations. Once students study and enact the scripts, they role play a tutoring situation in a less structured format, which leaves more to the students' invention and judgment. Performance in role playing can be critiqued and discussed by the teacher and by other classmates until the roles and expectations associated with peer tutoring are well established.

Short of a full-fledged peer tutoring program teachers may wish to train students to use specific rating scales, checklists, or related devices to assess their peers' performance. Students would not necessarily have to tutor peers who received low evaluations. They could, however, be trained to use instruments reliably and to make written or oral comments on particular types of behaviors or qualities.

Providing Feedback to the Class as a Whole

Feedback to the class as a whole may take a variety of forms and serve a variety of purposes, including:

1. To share with the class exemplary responses to test questions, e.g., a student may have written an essay that exemplifies desired qualities of insight and expression;
2. To be sure that students know the correct or best response to each item in a test, and the reasons why alternative responses are wrong or less adequate;
3. To identify learning areas in which a large number of students did exceptionally poorly, and in which special instructional attention needs to be given;
4. To identify learning areas in which a large number of students did exceptionally well, and in which praise or congratulations are called for, or in which opportunities for more advanced or independent work are to be provided; and
5. To permit students to interpret their scores in relation to scores obtained by others who took the test.

This last purpose, permitting students to view their performance in relation to that of others, may occasionally lead to inappropriate student competition. Many teachers complain that students are more interested in finding out where they stand in reference to classmates than in determining what they have or have not learned.

For this reason, information on the performance of a class as a whole (e.g., the average score; the range of scores, etc.) probably is not in itself particularly useful as an aid to learning. This is not to suggest that students should be denied information on the test performance of a class generally. It is only to point out that the main value of classroom tests to students is to confirm or reveal where they stand in relation to desired learning outcomes, not to show where they stand in relation to other students.

REFERENCES

Chapter 11

Bush, J.D., Bittner, J.R. & Brooks, W.D. The effect of the video-tape recorder on levels of anxiety, exhibitionism, and reticence. *Speech Teacher*, 1972, *21*, 127-130.

Niedermeyer, F.C. A model for the development or selection of school-based tutorial systems. In V.C. Allen (Ed.), *Children as teachers: Theory and research on tutoring*. New York: Academic Press, 1976

RELATED RESOURCES

Brophy, J. Teacher praise: A functional analysis. *Review of Educational Research*, 1981, *51* (1), 5-32.

Krumboltz, J.D. Tests and guidance: What students need. In W.B. Schrader (Ed.) *Measurement, guidance, and program improvement*. San Francisco, CA: Jossey-Bass, 1982.

Stewart, L. & White, M. Teacher comments, letter grades, and student performance: What do we really know? *Journal of Educational Psychology*, 1976, *68*, 488-500.

Ware, B. What rewards do students want? *Phi Delta Kappan*, 1978, *59*, 355-356.

SBB11

CHAPTER 12

USING TEST INFORMATION AS A GUIDE TO INSTRUCTIONAL DECISIONS

Introduction

Test information not only is useful to students. It also is useful to teachers as an aid in instructional decision making. Results from well-designed classroom tests carry important implications for teaching.

This chapter has three parts. The first offers a discussion of issues to consider when setting performance standards on tests. It is suggested that performance standards provide the basis for determining whether test results convey good news, bad news, or something in between.

The second presents guidelines for using test results to make instructional decisions about the **class generally**. The third part furnishes guidelines for making information-based decisions about **individual students** within the class. In both parts two and three common patterns of test performance are identified, and various options for responding to these patterns are discussed.

The chapter focuses on tests that are to be used to assess students' learning progress during a course, as distinct from tests given prior to instruction or tests given at the end of a semester or year.

Setting Performance Standards for a Test

Standards need to be established against which to evaluate test results. Standards indicate the degree of quality, proficiency, accuracy, speed, etc. that students must achieve in order to demonstrate that a learning outcome, or set of outcomes, has been attained. Without standards of acceptable test performance, it is difficult to know what to make of test findings.

Three questions need to be addressed when establishing performance standards: (1) How high should these standards be? (2) To whom should the standards apply? (3) Should standards be set in relation to subtests or a test as a whole?

With respect to the first question, about the height of standards, both the importance of an outcome and past evidence on students' outcome attainment are factors to consider in establishing an appropriate level of expectation. For outcomes that are viewed as essential for subsequent learning or for success in later life, particularly demanding performance standards often are set. For example, students in an English class may be expected to demonstrate that they can consistently write sentences with no errors in grammar and punctuation, whereas these same students may be permitted to make several mistakes when asked to identify the authors of particular literary selections, or the type of rhyme patterns used in various poems. Weaknesses in some learning areas may be more tolerable than weaknesses in other areas.

Information on the achievement levels of students in previous years who studied the same material also can be helpful in establishing appropriate standards. For example, if, over a five year period, a physics teacher found that most diligent students were capable of achieving a score of 85 percent correct on a test over principles of thermodynamics, a score of 85 percent might be considered as a "mastery" standard on this test. That is, students would be required to achieve this score to show that they had achieved mastery over the learning outcome(s) expected from the unit on thermodynamics. Information on patterns of student performance over several years, or across many different classrooms, may be helpful in determining a reasonable level of expectation on a test.

Answers to the question, "How high should standards be?" may be affected by answers to another question, "To whom should a given standard apply?" Standards typically are set in relation to all students in a class, but teachers who adopt a more personalized model of instruction may wish to establish different standards for students with different learning backgrounds or aptitude. For example, members of an accelerated learning group in a class may be expected to score 90 percent correct on a particular test, whereas other students may be expected to score 75 or 80 percent correct. In highly heterogeneous classes, varying standards by type of student may be appropriate.

The last issue to address when constructing performance standards concerns the aspect of the test in relation to which standards are to be set. A separate standard may be set for each subtest, or an overall standard may be set for a test as a whole. Standards for individual subtests are worth considering if the subtests are based on large enough samples of behavior to produce reliable scores. As suggested in chapter 11, the more numerous the subtests the smaller is the sample of behavior on which subscores are based. So long as confidence can be placed in subscores, it is reasonable to establish standards against which these subscores can be interpreted. This is particularly desirable when subtests cover distinctly different categories of learning outcomes, e.g., outcomes requiring students to remember facts vs. those requiring students to use principles.

Developing appropriate performance standards is thus a complex task. It involves considerable judgment on the part of teachers, and is influenced by a number of factors, including the nature of the outcomes to be assessed, evidence on the students' outcome attainment in previous years, the instructional model a teacher has adopted, and the composition of the test for which standards are to be formulated. The guidelines presented in the remainder of the chapter rest on the assumption that defensible performance standards for tests have been established.

Using Test Information to Guide Decisions About the Class As A Whole

Guidelines are presented in this section for dealing with evidence that the class has failed to meet established performance standards, or has exceeded performance standards. The ideas contained in this section build on ideas presented in Chapter 7: "Matching Teaching to Expected Learning Outcomes."

Responding to Evidence of Low Achievement

Reasons why students sometimes find it difficult to remember, use, or refine content are discussed in the paragraphs that follow. Suggestions for dealing with various causes of poor performance also are provided.

Condition 1: Students Have Difficulty Remembering Content

One possible cause:
Students did not have a clear idea about what they were supposed to remember.

In most courses, a good deal more material is introduced than is developed. Dozens of facts and concepts, for example, may be conveyed in an assigned chapter of a textbook, but only a few of these may be considered important enough to address explicitly in class. Likewise, an instructional film generally presents a great deal of information, only a portion of which is referred to in class discussion following the film.

One reason students may score poorly on items requiring them to remember content is that they don't know which of the large number of facts, concepts, principles, or procedures that have been covered are to be retained. This generally is not a problem for ideas or skills that are given explicit instructional attention. It is often a problem with background or supplementary information that is made available but not actively worked with. For information of this kind students may have little basis for choosing one element of knowledge to remember over another.

One solution to this problem is to exclude from tests any questions that require students to remember information that was not highlighted in some way in instructional presentations. In many courses, however, this would violate the purpose of independent reading assignments, field trips, or other activities through which information is presented that may relate only indirectly to the focus of classroom work. Teachers need to clarify for students how background or supplementary information will be sampled on tests, and from which sources, e.g., text, films, computer programs, etc.

Another Possible Cause:
Students have not made connections between material to be remembered and previous material learned, or personal experience.

Creating strong bridges for students between new material and familiar material helps them understand and remember what is presented. This is a rather obvious point, but if students do not remember content it may be because they have been unable to place it in a meaningful frame of reference. Teachers generally try to review previous material as a foundation for new learning, or link new learning to situations that have personal meaning to students. For example, when students are studying historical events, teachers commonly draw parallels between issues confronted in the past and issues found today. If student performance on remember-level tasks is poor, teachers may wish to make a greater effort to couple new ideas to familiar ones.

A Third Possible Cause:
Students did not have sufficient opportunity to verbalize or in other ways express the content to be remembered.

When students have a chance to present facts to classmates, restate or paraphrase definitions of concepts or principles, or describe, outline, or diagram procedures, they are more likely to remember them. If students' memory of content is weak, it may be because they lacked opportunity to express content as it was acquired.

There obviously are many other explanations of why students sometimes find it difficult to remember content, including failure to study for an examination or do a key homework assignment. The three causes addressed in the preceding paragraphs, however, are relatively common and should be looked to when thinking about ways to help students improve their learning.

Condition 2: Students Have Difficulty Using Content

One Possible Cause:
Students do not know content well enough to use it.

Performance at the use-level commonly depends on the successful acquisition and recall of content. Students who don't remember procedures for operating a word processor, or remember them incorrectly, for example, will have a hard time running a word processor. As other examples, students who forget or confuse a defining attribute of a concept may have trouble differentiating between examples and non examples of the concept, just as students who can't recall key formulas may be unable to solve a math problem. Difficulty in using content often results from an inability to remember content, or from a failure to acquire relevant content to begin with.

A single test occasionally provides information on both "use-level" learning goals and "remember-level" objectives intended to support them. On a unit test in chemistry dealing with atomic structure, for example, one section of the test may call upon students to paraphrase the laws of "conversion of mass," "definite proportions," and "multiple proportions." Another section may present problems that require the application of these laws, e.g., showing how the law of conversion of mass can be used to explain the changes that take place when a candle burns. (This would call for performance at the use level so long as students had not studied previously the burning candle example.) If a class scores low on use level (problem solving) tasks, teachers might check performance on the items at the remember-level. To the extent that students had difficulty remembering scientific laws, their failure on the problem solving items may be due to a lack of knowledge of the laws rather than an inability to apply them per se.

When test information does not contain information on students' knowledge of the content underlying a skill, simple learning probes may be designed to elicit this information. For example, if students consistently have trouble spiking a volleyball, a teacher may ask students to describe the steps that should be followed to execute a successful spike. If students are unable to recall these steps, or their recollections reveal misunderstandings, the teacher would have important clues about the origins of the problem in spiking.

Insofar as a weak knowledge base is the source of low scores on items requiring content application, teachers must either review relevant content or reteach it. Reviewing content is appropriate when the problem appears to be largely one of forgetting material. Reteaching is called for when students have fundamental misconceptions about the content they are to use.

Another Possible Cause:

Students have not had sufficient practice or feedback in using content.

If there is evidence that students know the concepts, principles, or procedures underlying a skill, but are unable to carry out the skill, the chances are that insufficient opportunity has been provided for students to use the skill and/or receive feedback on the quality, accuracy, speed, etc., of their performance. For example, if students know how one is supposed to change a sparkplug, but have difficulty in doing so, more practice in changing plugs, accompanied by specific feedback, probably is in order.

In some cases where practice in skill application is needed, there may be a need to structure practice more systematically rather than simply allot more time to practice. Usually, practice that is distributed over several time periods is more productive than a single, intensive practice session. It also is important to provide opportunities to practice in each type of situation in which students will be expected to perform on a test. For example, if students are expected to apply critical reading techniques to different types of reading material, e.g., newspaper editorials, scientific reports, and magazine advertisements, practice in reading each kind of material should be provided during instruction.

As the extent and nature of practice is modified, the value of modifying feedback procedures also may be considered. More specific feedback may be appropriate, accompanied by active demonstrations of how individual elements of performance might be improved.

Other plausible causes of difficulty in using content could be identified. One possibility that always should be investigated is that of establishing criteria for acceptable performance that are too high — although a more common occurrence is to set standards too low. However, when teachers are working in a new learning area, or with students whose backgrounds and abilities are different from those with whom they are accustomed to working, it often is difficult to establish performance criteria that are appropriate. In such cases, it often is helpful to talk with colleagues about reasonable criteria, or to defer setting specific criteria until opportunities for observing students have accumulated.

Condition 3: Students Have Difficulty Refining Content

Before identifying likely causes of poor performance at the refine level, it is important to note that teachers expect students to refine concepts, principles, or procedures less fequently than they expect them to remember and use them. Also, teachers are more likely to encourage students to modify or extend content than require them to do so. For example, students may receive instruction in techniques for photographing scenes of nature under differing conditions of light. The teacher may provide opportunities and reinforcement for students to build upon and vary the techniques presented to obtain more striking or lucid photographs. Chances are, however, that the teacher would not make such improvements a criterion of acceptable performance. Most teachers probably would be pleased if students could make effective use of established techniques, and would judge performance accordingly.

If performance at the refine-level is expected, but not achieved, there are at least two likely explanations:

One Possible Cause:

Students have not learned to work effectively at the use-level.

It is hard to make meaningful improvements in content without first having developed skill in applying it in established ways. There is an old adage that "discovery comes to the prepared mind." It no doubt is also true that extensions and modifications in current ideas and practices are most often made by those with wide experience in using them. If students are unable to find ways of improving established content, it may be because they lack depth of experience in applying it.

Another Possible Cause:

Students have not encountered situations that expose limitations in established content.

Unless students see a discrepancy between the requirements of a task and the tools they have to respond to it, there will be no motive for adapting tools or generating new ones. For example, there would be no reason to sharpen the definition of a concept if it provided an adequate basis for classifying all examples and nonexamples of the concept a teacher presented. If instructional presentations or learning activities have not brought into focus shortcomings in concepts, principles, or procedures, there is only a small chance that students will independently create improvements in them.

A Third Possible Cause:

Students have not received instruction in how to think divergently.

To work at the refine-level is to take intellectual risks. It is to create original responses, issues and solutions that may go against convention. It is to think imaginatively and independently. Without guidance in and reinforcement for this kind of divergent thinking, many students will never attempt it, and few will succeed in it.

* * * * * * * * * *

The common conditions of low achievement that have been identified in the paragraphs above and their likely causes are outlined in Table 12-1.

TABLE 12-1

Common Conditions of Low Class Achievement and Their Possible Causes

Condition	Possible Causes
1. Students have difficulty remembering content	1. Students did not have a clear idea about what they were supposed to remember.
	2. Students have not made connections between material to be remembered and previous material learned, or personal experience.
	3. Students did not have sufficient opportunity to verbalize or in other ways to express the content to be remembered.
2. Students have difficulty using content	1. Students do not know content well enough to use it.
	2. Students have not had sufficient practice and feedback in using content.
3. Students have difficulty refining content	1. Students have not learned to work effectively at the use-level.
	2. Students have not encountered situations that expose limitations in established content.
	3. Students have not received instruction in how to think divergently.

SBB12

Responding to Evidence of High Achievement

When students do exceptionally well on tests, here are some options that may be considered:

- Do nothing. Be pleased at the high levels of achievement.
- Provide exploratory or enrichment opportunities, e.g., read several additional articles on a particular topic; perform an experiment to confirm a scientific principle rather than simply reading about the principle, invite representatives from the community to debate political issues being studied in class, etc.
- Investigate the possibility that learning expectations have been set too low, and that students are capable of attaining more complex and sophisticated learning outcomes. Toward this end, teachers may ask students directly whether the work is too easy; they may ask colleagues who are teaching these students about the expectations they hold for student achievement; or they may consider the achievement levels of similar students who they have taught in the past.
- Investigate the possibility that the pace of instruction might be accelerated. Teachers may simply ask students whether the pace is too slow, or they might accelerate the pace and monitor closely the effect of the change.

Using Test Information to Guide Decisions About Individual Students

Responding to evidence that an individual consistently fails to meet criteria of acceptable performance on tests, or consistently attains desired learning goals at a faster rate than the majority of his or her classmates, is more complex than responding to evidence on students' performance generally. This is so for two reasons. One is that the factors explaining an individual's test performance often are more personal and idiosyncratic than the factors underlying the performance of students as a group. Individual learning styles or unusual talents need to be looked to, as well as the possibility of poor study habits, failure to study or weak instruction. Another is that time and resources needed to meet individual needs invariably are limited.

In this section, general suggestions for working with exceptionally low or high performing students are offered. These suggestions include options for diagnosing the nature of a student's exceptional performance, and for matching instruction to his or her learning background and characteristics. The focus is on students whose patterns of performance are consistently different from the norm, rather than on students whose performance is only occasionally unusually low or high.

Working with Low-Performing Students

1. Diagnosis

Teachers need to determine whether their students' poor performance is due primarily to: (1) a lack of prerequisites needed to succeed in the course, e.g., a physics student hasn't mastered the basic skills of mathematics; (2) a mismatch between a student's learning characteristics, on one hand, and teaching methods or the learning environment, on the other, e.g., a student accustomed to solitary learning is continually required to work in groups; or (3) some other cause, such as a high absence rate, disruptions in the home, etc.

(a) Determining whether poor performance is based on a lack of prerequisites. Teachers often identify individuals who lack basic knowledge and skills through interaction with students in the first few days or weeks of a class. Teachers also may use one or more of the following means to obtain information on an individual's learning deficits:

- administer a test of the basic knowledge and skills students need to meet course expectations. This could be a self-designed test: a test accompanying a textbook series or other instructional program; a district- or state-developed test; or one of the standardized tests of basic skills used on a national level;
- examine a student's permanent record for relatively recent evidence on basic skills achievement and level of preparation in the learning area in question (either recent test scores or information on the learning outcomes achieved in previous grades);
- refer the student to a counselor or assessment specialist for formal diagnosis of learning difficulties;
- talk with other teachers about the student's level of academic preparation.

The choice of one or more of these diagnostic approaches depends in part on policies and procedures established in individual districts, schools, or departments. In districts that give placement tests, or competency exams at the end of eighth grade, for example, information on students' basic skills attainment should be readily available. As another example, in schools that require teachers to describe students' end-of-course achievement level in reference to specific learning goals, a student's permanent record may indicate whether particular prerequisites for a course have been met. Finally, in schools that employ specialists to diagnose and remediate students' basic skill deficiencies, the classroom teacher's role in this regard typically is to refer low performing students to the specialist rather than to carry out systematic diagnosis on his or her own.

(b) Identifying a student's learning characteristics. An individual's poor academic performance may be due not to a lack of basic knowledge and skills, but to a mismatch between his or her learning characteristics and the teaching methods used for the class generally, or the approach taken to classroom management.

Although there are many different ways of classifying and identifying learning characteristics, most rest on theory and research on either **cognitive learning styles** (the characteristic patterns by which an individual receives and acts upon information), or **affective learning styles** (the characteristic patterns by which an individual is stimulated and motivated to learn or persevere in learning). Some students may be highly dependent on the teacher or highly susceptible to peer group pressure, for example, while other learners may be relatively self-directed. A number of resources for assessing learning characteristics may be found in Student Learning Styles: Diagnosing and Prescribing Programs, which is listed at the end of the chapter.

(c) Identifying other causes of poor performance. There obviously are many other reasons why a particular student may perform poorly, such as physical sickness, emotional problems, financial or family responsibilities, a history of failure in school, and resultant fear of academic demands. In cases like these, a personal conference with the student often aids in clarifying the nature of the problem. This is not to suggest that teachers should assume the role of a psychotherapist or social worker, but only that a discussion with a student may provide insights about the source of low achievement and a basis for dealing with it. Enlisting the assistance of counselors and parents in this process often is highly constructive.

2. Personalizing Instruction in View of Diagnostic Information

Diagnostic information of the kind referred to above can provide a basis for personalizing instructional methods and settings to improve the learning of low performing students. The degree to which teachers personalize instruction, however, depends a great deal on the instructional resources available to them, and the time they are able to devote to instructional planning. In this handbook, it is assumed that teachers will make some effort to adapt instruction to students who are not benefiting from approaches used for the class generally. However, it is recognized that the extent and form of a teacher's effort to personalize instruction will vary considerably according to the context in which the teacher is working.

(a) Overcoming deficiencies in basic knowledge and skills. Low performing students whose problems stem from lack of prerequisite skills require remedial instruction. In some schools special remedial courses are available, or educational specialists or instructional aides may be on hand to tutor students with learning deficiencies.

In many situations, however, the burden of remediation falls entirely on the classroom teacher. To the extent feasible, the teacher may work with individual students after school, or arrange for the student to have a peer tutor. Teachers also may wish to consult colleagues who teach in lower grades to obtain textbooks or individualized programs appropriate for poorly prepared students.

(b) Tailoring instructional methods and environments to a student's learning characteristics. There are as many ways to personalize instructional methods and environments as there are frameworks for classifying learning characteristics. One widely known model for matching instructional environments to individual learning traits is based on David Hunt's studies of students' "conceptual levels." Hunt's work is summarized briefly below as an illustration of the kind of matching that can be done between teaching approaches and student needs.

According to Hunt, students operate at different conceptual levels. A conceptual level is not the same as an achievement level. It is an indication of how a student learns rather than what he or she has learned. Hunt has identified three broad conceptual levels. The lowest, Stage A, is characterized by concrete and impulsive thinking with low tolerance for frustration. The second, Stage B, is characterized by conformity in thought, dependence on authority, and concern with "the right way" to do things. Students who operate at Stage B also tend to see solutions to problems in terms of black and white alternatives. At the highest stage of conceptual development (Stage C), students are inquiring, questioning, and intellectually self-assertive. They generate and consider a variety of alternative solutions to problems. Hunt maintains that the higher a student's conceptual level the less overt structure he or she needs from the learning environment. In Table 12-2, teaching practices that Hunt has found appropriate for students with differing needs for structure are identified.

(c) Dealing with special problems. Some students fail to achieve desired goals because of psychological, interpersonal, medical, or family problems that are unrelated to learning styles or past learning. Since there are countless numbers of these problems that might detract seriously from a student's school work, a universal prescription for dealing with such problems hardly can be defended. For example, a teacher may find that a phone call to a student's parents indicating to them that their son or daughter is often too sleepy to participate effectively in class leads to the establishment of a curfew for the student and a subsequent increase in the student's state of alertness. However, a teacher may also find that a student dozes in class because of a late-night job, which he or she has taken with parental encouragement. In this case, the teacher

TABLE 12-2

A Sample of Teaching Approaches Appropriate for Students with Differing Needs for Structure

Students Who Require A Great Deal of Structure

- Give specific guidelines and instructions (step-by-step), even make a chart of the steps.
- Make goals and deadlines short and definite.
- Provide a variety of activities during the period, incorporating some physical movement whenever possible.
- Make positive comments about their attempts; give immediate feedback on each step; give much assurance and attention; and praise them.

Students Who Require Some Structure

- Arrange students initially in rows and gradually get them working in pairs.
- Help them to know what to do each day.
- Provide non-threatening situations where they have to risk an opinion.
- Provide opportunities for choice and decision-making as they appear ready for them. Push them gently into situations where they have to make decisions and take responsibility.

Students Who Require Little Structure

- Allow them to select their own seats.
- Give them several topics from which to choose.
- Set weekly (or longer) assignments and allow students to make up their own timetables.
- Have them work in groups with the teacher serving as a resource person.

(Adapted from Hunt, 1979, pp. 36-37)

may decide to arrange a conference with the student's counselor and parents to explore the possibility of modifying work hours or changing jobs. A teacher's experience and judgment is essential in clarifying the specific nature of student problems such as these, and in working out plans to overcome them.

Working with High Performing Students

When an individual student consistently scores exceptionally high on tests or achieves desired goals at a much faster rate than anticipated, a number of options may be considered, including:

- placing the student in an accelerated or honors class (this would normally involve a conference with the student and his parents, a counselor, and the teacher of the advanced class);
- giving the student extra-credit assignments for enrichment purposes;
- asking the student to serve as a peer tutor or a tutor of younger children; or
- developing an independent study program for the student to permit more sophisticated and extensive exploration of the topics the class as a whole is addressing, and/or the pursuit of more advanced topics.

The last option, independent study programs, often involves creating "learning contracts" with students. The level of detail and prescriptiveness of a contract depends on a student's age and maturity, but typically contracts include:

- expected learning outcomes, stated in language appropriate for students (students may have a hand in selecting or generating expected outcomes);
- an indication of how progress toward attaining learning goals will be assessed, and how achievement at the end of the project will be measured;
- a description of the resources students should draw upon to carry out the project, e.g., authorities in the community, statistical data from government reports, journal articles, texts, etc.
- a set of checkpoints specifying when the teacher and a student are to meet to confirm and reinforce progress and to identify and resolve problems;
- a set of deadlines for completing various aspects of the project;
- an indication of the contribution that work in the unit will make to a student's grades for a quarter or course.

REFERENCES

Chapter 12

Hunt, D.E. Learning style and student needs: An introduction to conceptual level. In Student learning styles: Diagnosing and prescribing programs. Reston, Va.: National Association of Secondary School Principals, 1979.

RELATED RESOURCES

Block, J.H. Student learning and the setting of mastery performance standards. Educational Horizons, 1972, 50, p. 183-191.

Klausmeir, H.K., Lipham, J.M. & Daresh, J.C. The renewal and improvement of secondary education: Concepts and practices. Lanham, Md.: University Press of America, 1983.

Mager, R.F. & Pipe, P. Analyzing performance problems. Belmont, Ca.: Fearon Publishers, 1970.

Okey, J. Altering teacher and pupil behavior with mastery teaching. School Science and Mathematics, 74, no. 6, 1974, 530-535.

Ryan, D.W. & Schmidt, M. Mastery learning: Theory, research and implementation. Ontario, Canada: Ministry of Education, 1979.

Wanous, D. & Mehrens, W. Helping teachers use information: The data box approach. Measurement in Education, 12, 1981, 1-10.

SBB12

CHAPTER 13

USING TEST INFORMATION IN GRADING

Introduction

In a recent study of high school teacher's testing practices (Fielding & Schalock, 1983), more teachers reported using test information as a basis for grading than for any other purpose. In most classrooms, there is a close relationship between the test scores students receive and the grades they are assigned.

The approach to grading developed in this chapter rests on the assumption that grades should reflect students' achievement of the learning outcomes desired from instruction. Factors other than achievement, like effort exerted in learning, or study habits are not viewed as pertinent in assigning grades. The position is taken that grades should be based on evidence of students' attainment of established learning goals.

The chapter is organized around five broad steps that need to be taken when developing a goal-based approach to grading:

(1) Clarify the learning goals students are expected to attain by the end of a grading period, semester, or year, and on which grades will be based;

(2) If different learning goals are established for different groups of students, clarify the implications of these differences for grading;

(3) Identify the sources of evidence on goal attainment to be used in calculating a grade;

(4) Create a procedure for translating evidence from multiple measures over varied goals into a single grade; and

(5) Relate grading standards to standards of acceptable performance that have been established to guide instructional decisions.

Although it is not always feasible to carry out each of these steps in the beginning of a course or semester, the steps should be completed as early as possible in a grading period. This will help assure that students have a clear idea of the connection between the learning to be accomplished, learning demonstrated, and grades.

Clarify the Learning Goals Students are Expected to Attain by the End of a Grading Period, Semester, or Year, and on which Grades will be Based

In a goal-based approach to grading students need to know which learning goals they are expected to accomplish by the end of each instructional period for which grades are to be assigned. For example, if teachers plan on reporting grades to students and parents every ten weeks, then students should be clear about the learning goals to be accomplished by the end of each ten week period.

Since teachers generally organize instruction around learning goals established for a unit rather than a 10-week grading period, the connection between outcomes expected from units and those expected by the end of a grading period needs to be made clear. Some teachers in some subject areas may decide that the learning outcomes appropriate for a grading period are nothing more than the sum of all learning outcomes expected from the individual units taught during the period. If, for example, three units are taught, each guided by six learning goals, then students might be expected to accomplish a total of 18 goals by the end of the grading period.

In some cases, however, more cumulative, long-term learning outcomes may be expected by the end of a grading period than by the end of particular instructional units. For example, an English teacher may teach a unit on outlining in preparation for two units on essay writing. While the teacher may desire students to develop outlining skills as a short-term goal, the more important learning goal is the ability to write well-organized essays. The teacher may wish to base students' grades for a 10-week period on the quality of essays they write rather than on the outlines they produce, even though outlining was emphasized during an introductory instructional unit.

In cases like this, students need to know that their performance in relation to "supporting goals," i.e., goals that are to be attained as a basis for attaining larger and more important outcomes, is not to be considered in calculating a grade, or is to be weighed less heavily in assigning grades than their performance in relation to long-term goals.

The ideas discussed in the preceding paragraphs also apply when computing grades for a course as a whole. Course grades may reflect students' achievement of the learning goals expected from each grading period, (calculated, for example, by averaging grades received for each individual grading period), or they may reflect more closely students' achievement of outcomes that are cumulative in nature, i.e., those that rest upon knowledge and skills developed over two or more grading periods. An example of such an outcome would be the ability to design a viable and cost-efficient solar energy heating system for a new building being constructed in the community.

The relative emphasis to be placed on students' achievement of short-term vs. long-term learning goals in assigning grades needs to be clarified as early in a course as possible.

If Different Learning Goals are Established for Different Groups of Students, Clarify the Implications of These Differences for Grading

As indicated in Chapter 6, some teachers tailor learning goals to the backgrounds and abilities of individuals or groups of students. Students who enter a class with an extensive background in a learning area, for example, may be expected to reach more advanced goals than students who begin a course with limited backgrounds.

When goals vary by individuals or groups of students, so, too, may grading standards. High grades, for example, may be given to a low-ability student who attains expected learning outcomes that are less sophisticated than the outcomes set for academically gifted students.

Whether or not a teacher varies grading standards to accommodate differences among students obviously depends in large part on policies established by a school or district. In some schools, teachers may be expected to vary learning expectations and grading standards to accommodate different ability groups within classes that are not "tracked," that is, classes which contain students of widely differing abilities. By contrast, in classes that are tracked, or are so specialized that they are selected only by students with particular learning backgrounds, e.g., calculus or physics, teachers might be expected to hold all students accountable to the same set of standards.

To the extent that school policies require, encourage, or permit teachers to vary expected outcomes and grading standards according to student aptitudes, it is essential that a plan for accomplishing this be developed with care and made explicit to students, parents, and administrators.

Identify the Sources of Evidence on Goal Attainment to be Used in Calculating a Grade

Teachers commonly use more than one procedure in assessing student achievement. Quizzes, homework assignments, projects and tests are among the most commonly used measures of student learning. Teachers do not necessarily use evidence from each of these measures, however, in calculating a grade. For example, a teacher may collect first drafts of reports or other products to determine what additional assistance a student needs to produce a high-quality piece of work. In most cases, a teacher would not assign a grade to this first draft, but would defer grading until the product is in final form. Similarly, teachers would give greater weight to more valid measures of student learning than less valid measures. Students need to know which sources of evidence on goal attainment are to be used in calculating their grade.

Develop a Procedure for Translating Evidence from Multiple Measures Over Varied Goals into a Single Grade

When evidence from multiple measures over varied goals is to be used in calculating a grade, teachers need to be clear about the weight to be given to each measure over each goal. Several different procedures for translating evidence on goal attainment into a grade are illustrated below.

Example 1:

Convert all assessment results into point values that represent the relative importance of each learning outcome that has been assessed, and the nature of the measure that was used to assess outcome attainment. Total point values for a marking period can be readily translated into grades, as illustrated in Table 13-2.

Example 2:

In this example, a distinction is made between "core goals," which are viewed as especially important, and "other goals," which are considered worthwhile, but less essential. Grading standards in this example are more demanding with respect to core goals than other goals, as illustrated below.

Grade		Core Goals		Other Goals
A	=	An average of 90 percent correct on all tests or subtests over core goals	+	An average of 80 percent correct on all tests or subtests over other goals
B	=	An average of 80 percent correct on all tests or subtests	+	An average of 70 percent correct on all tests or subtests
C	=	An average of 70 percent correct on all tests or subtests	+	An average of 60 percent correct on all tests or subtests

TABLE 13-2

A Goal-Based Point System for Grading

Goal	Relative Importance	Total Possible Points				
		Quizzes	Projects	Tests		
Goal A	High	7	20	30	=	57
Goal B	High	5	20	34	=	59
Goal C	Moderate	4	15	20	=	39
Goal D	Moderate	4	15	20	=	39
Goal E	Low	2	—	11	=	13
Goal F	Low	3	—	10	=	13
		25	70	125	=	220

Grading Standards

A = 190 - 220
B = 160 - 189
C = 125 - 159
D = 90 - 124
E = below 90

SBB13

Example 3:

Use a simple mastery-nonmastery standard for evaluating student performance in relation to individual goals, and base grades on the number of goals over which mastery has been demonstrated, e.g.:

Number of Goals Attained	Grade Assigned
11-12	A
9-10	B
7-8	C
5-6	D
Fewer than 5	F

Relate Grading Standards to Standards Used to Guide Instructional Decisions

As described in Chapter 12, it is desirable to establish standards of acceptable performance on tests as a basis for interpreting test results. However, performance standards used to guide instructional decisions are not the same as grading standards. A student may achieve a learning outcome at the level of mastery called for in a standard established to guide instructional decisions, but whether this warrants a grade of "A," "B," or "C" is another issue. To one teacher, "acceptable performance" on a test of outcome attainment may warrant a grade of "B," whereas another may decide that it is worth only a "C." The important point, however, is that the relationship between standards of acceptable performance, or standards of "mastery," and grades needs to be made clear to students.

A school's policy on grading may affect the relationship a teacher establishes between mastery standards and grading standards. In a school that has adopted a traditional grading model, for example, in which students are expected to compete against each other for a limited number of "A's," and "B's," grades must be distributed in a class according to a curve. It is expected that a small group, perhaps 15 percent of the class, will be assigned an "A," a somewhat larger percentage will be assigned a "B," and so forth. Under such a model, students' grades are not simply based on where they stand with respect to a mastery standard, but where they stand in relation to each other. The connections and distinctions a teacher makes between standards guiding instructional decisions and standards guiding grade assignments may therefore depend in large part on grading policies that prevail in a school or district.

REFERENCES

CHAPTER 13

Fielding, G.D. & Schalock, H.D. A survey of high school teachers' testing practices. Monmouth, Or.: Teaching Research Division, Oregon State System of Higher Education, 1983.

RELATED RESOURCES

Bloom, B.S., Madaus, G.F. & Hastings, J.T. Evaluation to improve learning. New York: McGraw-Hill, 1981.

Gronlund, N.E. Improving marking and reporting in classroom instruction. New York: Macmillan, 1974.

Terwilliger, J.S. Assigning grades to students, Glenview, Ill.: Scott, Foresman, 1971.

Terwilliger, J.S. Assigning grades: Philosophical issues and practical recommendations. Journal of Research and Development in Education, 1977, 10 (3), 21-39.

SBB13